

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2002, 19:34:03 ; Search time 253 Seconds
(without alignments)
9168.214 Million cell updates/sec

Title: US-09-945-376-3
Perfect score: 1030
Sequence: 1 ctctatagggcagcggtg.....tgtatatatacaagggtctgg 1030

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
C 1	45	4.4	3548	24	AAD38882
2	42.4	4.1	6078	22	AAF31861
3	41	4.0	2614	19	AAVA0621
4	40.6	3.9	1038	21	AAC62796
5	40.6	3.9	1038	24	ABK17061
6	40.4	3.9	598	24	AAS15016
7	40.4	3.9	2096	21	AAC62810
8	40.4	3.9	2096	24	ABK17075
C 9	40.4	3.9	2611	24	ABA91255

10	40.2	3.9	411	21	AAC62780	Flower specific pr
11	40.2	3.9	411	24	ABK17045	Eucalyptus grandis
12	40	3.9	594	20	AAK35652	5' region of human
13	40	3.9	594	21	AAA75054	Upstream sequence
14	40	3.9	1680	20	AAV08328	Hyaluronate syntha
15	39.8	3.9	1224	21	AAZ51547	Seed-preferred pro
16	39.8	3.9	2385	21	AAC83331	PART-1 promoter re
17	39.8	3.9	2475	24	ABJ50287	Rat glyceraldehyde
18	39.8	3.9	2565	24	ABJ50285	Rat glyceraldehyde
19	39.8	3.9	3718	20	AAV72244	G. max SBP2 DNA 5'
20	39.8	3.9	3718	24	AAD38881	Alfalfa AP2 adapto
21	39.4	3.8	288	21	AAC62788	Isoflavone reducta
22	39.4	3.8	288	24	ABK17053	Eucalyptus grandis
23	39.4	3.8	515	21	AAC62816	MIF homologue codi
24	39.4	3.8	515	24	ABK17081	Eucalyptus grandis
C 25	2492	3.8	2867	19	AAV61948	T. thermophila del
26	39.4	3.8	2867	19	AAV68599	Lecithin-cholester
27	39	3.8	311	21	AAC62809	Flower specific pr
28	39	3.8	311	24	ABK17074	Eucalyptus grandis
29	39	3.8	516	24	AAS15015	Rice tubulin-like
30	39	3.8	927	21	AAC62779	Flower specific pr
31	39	3.8	927	24	ABK17044	Eucalyptus grandis
32	39	3.8	940	24	AAD38884	Alfalfa plastocyan
33	39	3.8	1126	21	AAC62802	Pollen specific pr
34	39	3.8	1126	24	ABK17067	Pinus radiata prom
C 35	39	3.8	1241	19	AAV18017	Pinus radiata cone
C 36	39	3.8	1242	18	AAT94308	PrFL1 gene promote
37	39	3.8	4526	20	AAV72245	G. max SBP1 DNA 5'
C 38	38.6	3.7	1196	22	AAI68306	Aspergillus oryzae
C 39	38.6	3.7	5031	22	AAF76068	Maize MADS-box gen
C 40	38.6	3.7	5031	22	AAF76455	Maize ZmMADS2 gene
C 41	38.4	3.7	382	21	AAC62789	Glyceraldehyde-3-p
C 42	38.4	3.7	382	24	ABK17054	Eucalyptus grandis
C 43	38.4	3.7	648	21	AAC62787	Isoflavone reducta
C 44	38.4	3.7	648	24	ABK17052	Eucalyptus grandis
C 45	38.4	3.7	921	21	AAZ51580	Maize seed-prefer

ALIGNMENTS

RESULT 1
AAD38882/c
ID AAD38882 standard; DNA; 3548 BP.
XX
AC AAD38882;
XX
DT 23-SEP-2002 (first entry)
DE Alfalfa plastocyanin gene coding fragment.
XX
KW Expression regulatory sequence; recombinant polypeptide; gene cloning;
KW alfalfa; plastocyanin; ds.
XX
OS Medicago sativa.
XX
PN WO200236786-A2.
XX
PD 10-MAY-2002.
XX
PF 31-OCT-2001; 2001WO-CA01532.
XX
PR 31-OCT-2000; 2000US-244214P.
XX
PA (MEDI-) MEDICAGO INC.
XX
PI Vezina L, D'aoust M, Arcand F, Bilodeau P;
XX
DR WPI; 2002-471503/50.
XX
PT Isolating and characterizing an expression regulatory sequence for
PT expressing recombinant polypeptides and/or RNA, comprises producing
PT oligonucleotide primers that amplify sequences upstream or downstream

PT of cDNAs
XX Example 2; Page 68-69; 74pp; English.
XX The invention relates to a method for isolating and characterising an
CC expression regulatory sequence for the expression of recombinant
CC polypeptides and/or RNA. The method comprising producing at least one
CC oligonucleotide primer from cDNAs of a cDNA library, where the
CC oligonucleotide primer allows amplification of genomic sequences upstream
CC or downstream of the cDNAs. The method is useful for isolating,
CC characterising and identifying a large number of known and unknown
CC promoters that are active under any desired environmental condition to
CC which a cell may be exposed and not just to the exemplified isolation of
CC promoters that are capable of expression in specific conditions. The
CC methods are also useful for cloning genes from any host, or from a
CC specific tissue with such host, from which a cDNA library may be
CC constructed; for the identification and isolation of analogous promoters,
CC and signal peptides and structural genes in several species of multicellular
CC and unicellular organisms and as a high throughput identification system
CC of candidate therapeutic targets. The promoter sequences may be used to
CC regulate the synthesis of polypeptides. The present sequence is alfalfa
CC plastocyanin gene coding fragment.
XX
SQ Sequence 3548 BP; 1112 A; 576 C; 727 G; 1133 T; 0 other;

Query Match 4.4%; Score 45; DB 24; Length 3548;
Best Local Similarity 90.6%; Pred. No. 0.0029;
Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 TTACTATAGGCGCGTGTGTCGACGCGCGGCTGTATGAGTGGGAC 54
|||||
Db 3529 TTACTATAGGCGCGTGTGTCGACGCGCGGCTGTATGAGTGTGCACC 3477
|||||

RESULT 2
AAF31861
ID AAF31861 standard; DNA; 6078 BP.
XX
AC AAF31861;
XX
DT 12-APR-2001 (first entry)
XX
DE Human KARP-1 DNA.
XX
KW Human; KARP-1; Ku86 autoantigen related protein; cancer;
KW immune deficiency disorder; biliary tract cancer; leucine zipper protein;
KW cytostatic; immunosuppressant; gene therapy; KARP-1 inhibitor; ds.
XX
OS Homo sapiens.
XX
PN US6171857-B1.
XX
PD 09-JAN-2001.
XX
PF 16-OCT-1998; 98US-0173914.
XX
PR 17-OCT-1997; 97US-0064557.
XX
PA (UYBR-) UNIV BROWN RES FOUND.
XX
PI Hendrickson EA;
XX
DR WPI; 2001-146208/15.
DR P-PSDB; AA866590.
XX
PT Novel nucleic acids encoding leucine zipper protein, KARP-1
PT polypeptide, useful for treating cancer and immune deficiency disorder
PT
XX
PS Claim 1; Column 51-56; 61pp; English.
XX
CC The present sequence is given in a specification relating to an isolated
CC Ku86 Autoantigen Related Protein (KARP-1) nucleic acid molecule. The

CC KARP-1 nucleic acid and KARP-1 protein are useful for the treatment
CC and/or diagnosis of diseases such as cancer and immune deficiency
CC disorders. They are useful in combination with a KARP-1 inhibitor that
CC inhibits double stranded DNA base repair. Inhibitors of KARP-1 are
CC useful in the diagnosis or treatment of conditions characterised
CC by the loss of KARP-1 activity and in the treatment of cancer,
CC e.g. biliary tract cancer.
XX
SQ Sequence 6078 BP; 1806 A; 1301 C; 1392 G; 1543 T; 36 other;

Query Match 4.1%; Score 42.4; DB 22; Length 6078;
Best Local Similarity 97.7%; Pred. No. 0.028;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTTACTATAGGCGCGTGTGTCGACGCGCGGCTGTATGAA 44
|||||
Db 9 CTTACTATAGGCGCGTGTGTCGACGCGCGGCTGTATGAA 52
|||||

RESULT 3
AAV40621
ID AAV40621 standard; DNA; 2614 BP.
XX
AC AAV40621;
XX
DT 26-OCT-1998 (first entry)
XX
DE Cotton promoter region from an expansin gene.
XX
KW Cotton fibre expansin promoter; transgenic; ss.
XX
OS Gossypium hirsutum
XX
PN WO9830698-A1.
XX
PD 16-JUL-1998.
XX
PF 07-JAN-1998; 98WO-US00151.
XX
PR 07-JAN-1997; 97US-0034914.
XX
PA (CALJ) CALGENE INC.
XX
PI Pear JR, Stalker DM;
XX
DR WPI; 1998-399143/34.
XX
PT New DNA containing the promoter of the cotton expansin gene - used
PT for tissue-selective expression of genes that alter cotton fibre
PT phenotype
XX
PS Claim 2; Fig 1a-d; 26pp; English.
XX
CC The present sequence represents an expansin promoter region from
CC Gossypium hirsutum. This has been sequenced and found to be
CC approximately 2200 bp in length, and is immediately 5' to the cotton
CC fibre expansin coding region. The promoter can be used in cotton
CC transformation by linking it to a gene of interest, thus enabling the
CC modification of transgenic cotton fibre cells.
XX
SQ Sequence 2614 BP; 892 A; 436 C; 401 G; 882 T; 3 other;

Query Match 4.0%; Score 41; DB 19; Length 2614;
Best Local Similarity 100.0%; Pred. No. 0.048;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTTACTATAGGCGCGTGTGTCGACGCGCGGCTGTAT 41
|||||
Db 83 CTTACTATAGGCGCGTGTGTCGACGCGCGGCTGTAT 123
|||||

RESULT 4
AAC62796

ID AAC62796 standard; DNA; 1038 BP.
XX AAC62796;
AC
XX 02-FEB-2001 (first entry)
DT
XX Senescence-like protein promoter coding sequence #1.
DE
XX Promoter; eucalyptus; pine; gene transcription; ds.
KW
XX Pinus radiata.
OS
XX WO200058474-A1.
PN
XX 05-OCT-2000.
PD
XX 24-FEB-2000; 2000WO-NZ00018.
PF
XX 25-MAR-1999; 99US-0276599.
PR
XX 30-JUL-1999; 99US-0146591.
PR
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
PA
XX Perera R, Rice SJ, Eagleton CK;
XX WPI; 2000-647236/62.
DR P-PSDB; AAB28136.
DR
XX Novel promoter sequences useful for modulating transcription of plant
XX DNA sequences of interest and production of polypeptides
XX Claim 1; Pages 57-58; 93pp; English.
XX The present invention relates to promoter sequences from eucalyptus
XX and pine. The present sequence is one such promoter. This sequence is
XX useful for modulating the transcription of DNA sequences of interest. The
XX sequences may also be used to tag or identify an organism or its
XX reproductive material.
XX
SQ Sequence 1038 BP; 298 A; 220 C; 233 G; 287 T; 0 other;
Query Match 3.9%; Score 40.6; DB 21; Length 1038;
Best Local Similarity 71.1%; Pred. No. 0.037;
Matches 69; Conservative 0; Mismatches 24; Indels 4; Gaps 1;
QY 2 TTACTATAGGCGACGCGTGTGTCGACGCCGCGGCTGGTATGAAGGTGGGAACCTCCTG 61
DB 4 TTACTATAGGCGACGCGTGTGTCGACGCCGCGGCTGGTATAA----TGAGAACATGATAAG 59
QY 62 ATGCATATATCTGCTGAGAGATACATCAATTCACA 98
DB 60 CTGTGTAATTCATGCTAGTCACCATAACTTTCTCA 96
RESULT 5
ABK17061
ID ABK17061 standard; cDNA; 1038 BP.
XX
XX ABK17061;
AC
XX 26-MAR-2002 (first entry)
DT
XX Pinus radiata promoter polynucleotide #20.
DE
XX Promoter; pine; leaf; root; flower; bud; meristem; xylogenesis;
KW temporally regulated promoter; Pinus radiata; Eucalyptus grandis; ss;
KW PCR primer.
XX
XX Pinus radiata.
OS
XX WO200198485-A1.
PN
XX

PD 27-DEC-2001.
XX
PF 20-JUN-2001; 2001WO-NZ00115.
XX
XX 20-JUN-2000; 2000US-0598401.
PR
XX 28-NOV-2000; 2000US-0724624.
PR
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS IND LTD.
PA
XX Perera R, Rice S, Eagleton C, Lasham A;
XX WPI; 2002-114583/15.
DR P-PSDB; AAU80754.
DR
XX Novel polynucleotide promoter sequences from Pine and Eucalyptus useful
XX for modifying expression of endogenous and/or heterologous
XX polynucleotides in transgenic plants
XX Claim 1; Page 73-74; 121pp; English.
XX The invention relates to isolated promoter sequences from Pinus radiata
XX and Eucalyptus grandis, comprising a sequence chosen from leaf-, root-,
XX flower-, pollen-, bud-, meristem-specific promoters or temporally
XX regulated promoters such as xylogenesis-specific promoters. The promoter
XX polypeptides and their related polynucleotides are useful in the
XX production of genetic constructs, used for modifying gene expression in a
XX target organism, in particular a plant. The method is useful for
XX modifying expression of a polynucleotide that comprises an intron
XX sequence, through removal of the intron sequence. The method is useful
XX for modifying growth and development of plants, and cellular responses to
XX external stimulus, such as environmental factors and disease pathogens.
XX The sequences are useful in genome and physical mapping, in positional
XX cloning of genes, in various assays to determine biological activity, to
XX raise antibodies, to isolate corresponding interacting proteins and other
XX compounds, and to quantitatively determine levels of interacting proteins
XX or other compounds. Sequences ABK17016-ABK17125 represent Pinus radiata
XX and Eucalyptus grandis polynucleotides and PCR primers used in the method
XX of the invention.
XX
SQ Sequence 1038 BP; 298 A; 220 C; 233 G; 287 T; 0 other;
Query Match 3.9%; Score 40.6; DB 24; Length 1038;
Best Local Similarity 71.1%; Pred. No. 0.037;
Matches 69; Conservative 0; Mismatches 24; Indels 4; Gaps 1;
QY 2 TTACTATAGGCGACGCGTGTGTCGACGCCGCGGCTGGTATGAAGGTGGGAACCTCCTG 61
DB 4 TTACTATAGGCGACGCGTGTGTCGACGCCGCGGCTGGTATAA----TGAGAACATGATAAG 59
QY 62 ATGCATATATCTGCTGAGAGATACATCAATTCACA 98
DB 60 CTGTGTAATTCATGCTAGTCACCATAACTTTCTCA 96
RESULT 6
AAS15016
ID AAS15016 standard; cDNA; 598 BP.
XX
XX AAS15016;
AC
XX 14-FEB-2002 (first entry)
DT
XX Rice anther-specific promoter #2.
DE
XX Rice; promoter; gene therapy; transcription; monocot; anther;
KW chromosome mapping; gene mapping; antisense technology;
KW plant genetic engineering; ss.
XX
XX Oryza sativa.
OS
XX WO200181606-A2.
PN
XX

CC production of genetic constructs, used for modifying gene expression in a
 CC target organism, in particular a plant. The method is useful for
 CC modifying expression of a polynucleotide that comprises an intron
 CC sequence, through removal of the intron sequence. The method is useful
 CC for modifying growth and development of plants, and cellular responses to
 CC external stimulus, such as environmental factors and disease pathogens.
 CC The sequences are useful in genome and physical mapping, in positional
 CC cloning of genes, in various assays to determine biological activity, to
 CC raise antibodies, to isolate corresponding interacting proteins and other
 CC compounds, and to quantitatively determine levels of interacting proteins
 CC or other compounds. Sequences ABK17016-ABK17125 represent Pinus radiata
 CC and Eucalyptus grandis polynucleotides and PCR primers used in the method
 CC of the invention.

XX
 SQ Sequence 2096 BP; 537 A; 535 C; 539 G; 485 T; 0 other;

Query Match 3.9%; Score 40.4; DB 24; Length 2096;
 Best Local Similarity 97.6%; Pred. No. 0.065;
 Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTACTATAGGCACGGCTGTCGACGCCCGGGCTGGTATGA 43
 ||||||||||||||||||||||||||||||||||||||||
 Db 3 TTACTATAGGCACGGCTGTCGACGCCCGGGCTGGTCTGA 44

RESULT 9

ABA91255/C
 ID ABA91255 standard; DNA; 2611 BP.

XX
 AC ABA91255;

XX
 DT 04-APR-2002 (first entry)

XX
 DE Arabidopsis BONSA11 (BON1) gene, involved in growth homeostasis.

XX
 KW BONSA11; BON1; phospholipid binding protein; growth; homeostasis;
 XX thermotolerance; transgenic plant; plant; gene; ds.

XX
 OS Arabidopsis thaliana.

XX
 FN WO200200697-A2.

XX
 PD 03-JAN-2002.

XX
 PF 25-JUN-2001; 2001WO-US20172.

XX
 PR 23-JUN-2000; 2000US-213863P.

XX
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX
 PI Hua J, Grisafi P, Fink GR;

XX
 DR WPI; 2002-139899/18.

XX
 PT New phospholipid binding proteins and nucleic acids, useful for
 PT modulating plant growth homeostasis, controlling cell expansion and
 PT cell division, or producing plants where larger fruits and increased
 PT biomass are desired -

XX
 PS Disclosure; Fig 8A; 78pp; English.

XX
 CC The present sequence is that of the BONSA11 (BON1) gene of
 CC Arabidopsis thaliana. The coding region of the BON1 gene is
 CC given in ABA91255. The BON1 gene permits wild-type Arabidopsis
 CC plants to maintain a relatively constant size over a wide range of
 CC temperatures. Thus, bon1 null mutants produce miniature fertile
 CC plants at 22 degrees C, but a wild-type phenotype at 28 degrees C.
 CC BON1 has a direct role in regulating cell expansion and cell
 CC division at temperatures lower than those at which Arabidopsis is
 CC normally grown. The BON1 protein (see ABA91255) contains a
 CC Ca2+-dependent phospholipid binding domain and is associated with
 CC the plasma membrane. BON1 belongs to the copine gene family, which
 CC is conserved from protozoa to humans. The invention is directed to

CC isolated BON1, BON2, BON3, BAP1 and BAL nucleic acids (see
 CC ABA91256-60), which encode proteins (see ABA91256-60) that are
 CC necessary for normal growth, controlling cell expansion and cell
 CC division, affecting the size and rate at which the plant grows when
 CC exposed to lower temperatures. Transgenic plants are provided that
 CC are smaller than the wild-type as a result of inhibition of BON1,
 CC BON2, BON3, BAP1 and/or BAL, especially angiosperms and gymnosperms,
 CC ornamental plants and turfgrass. Transgenic plants are also
 CC provided that are larger than the wild-type as a result of
 CC enhancement of BON1, BON2, BON3, BAP1 and/or BAL, especially crop
 CC plants and biomass plants. Modulation of these genes provides
 CC increased yield, or growth at a higher altitude or lower
 CC temperature.

XX
 SQ Sequence 2611 BP; 809 A; 429 C; 498 G; 866 T; 9 other;

Query Match 3.9%; Score 40.4; DB 24; Length 2611;
 Best Local Similarity 97.6%; Pred. No. 0.074;
 Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCACGGCTGTCGACGCCCGGGCTGGTATG 42
 ||||||||||||||||||||||||||||||||||||||||
 Db 2603 CTTACTATAGGCACGGCTGTCGACGCCCGGGCTGGTCTG 2562

RESULT 10

AAC62780
 ID AAC62780 standard; DNA; 411 BP.

XX
 AC AAC62780;

XX
 DT 02-FEB-2001 (first entry)

XX
 DE Flower specific promoter coding sequence #2.

XX
 KW Promoter; eucalyptus; pine; gene transcription; ds.

XX
 OS Eucalyptus grandis.

XX
 PN WO200058474-A1.

XX
 PD 05-OCT-2000.

XX
 PF 24-FEB-2000; 2000WO-NZ00018.

XX
 PR 25-MAR-1999; 99US-0276599.

XX
 PR 30-JUL-1999; 99US-0148591.

XX
 PA (GENE-) GENESIS RES & DEV CORP LTD.

XX
 PI (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX
 PI Perera R, Rice SJ, Eagleton CK;

XX
 DR WPI; 2000-647236/62.

XX
 PT Novel promoter sequences useful for modulating transcription of plant
 PT DNA sequences of interest and production of polypeptides -

XX
 PS Claim 1; Page 53; 93pp; English.

XX
 CC The present invention relates to promoter sequences from eucalyptus
 CC and pine. The present sequence is one such promoter. This sequence is
 CC useful for modulating the transcription of DNA sequences of interest. The
 CC sequences may also be used to tag or identify an organism or its
 CC reproductive material.

XX
 SQ Sequence 411 BP; 112 A; 118 C; 80 G; 101 T; 0 other;

Query Match 3.9%; Score 40.2; DB 21; Length 411;

Best Local Similarity 93.3%; Pred. No. 0.029;
 Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ACTATAGGCACGGCTGTCGACGCCCGGGCTGGTATGAAGGTG 48

Db 1 ACTATAGGACCGGTGGTCGACGGCCGGGCTGGTCTGAACGTG 45
|||||

RESULT 11

ABK17045
ID ABK17045 standard; cDNA; 411 BP.

XX AC ABK17045;

XX DT 26-MAR-2002 (first entry)

XX DE Eucalyptus grandis promoter polynucleotide #16.

XX KW Promoter; pine; leaf; root; flower; pollen; bud; meristem; xylogenesis;
KW temporally regulated promoter; Pinus radiata; Eucalyptus grandis; ss;
KW PCR primer.

XX OS Eucalyptus grandis.

XX PN WO200198485-A1.

XX PD 27-DEC-2001.

XX PF 20-JUN-2001; 2001WO-NZ00115.

XX PR 20-JUN-2000; 2000US-0598401.

XX PR 28-NOV-2000; 2000US-0724624.

XX PA (GENE-) GENESIS RES & DEV CORP LTD.

XX PA (FLET-) FLETCHER CHALLENGE FORESTS IND LTD.

XX PI Perera R, Rice S, Eagleton C, Lasham A;

XX DR WPI; 2002-114583/15.

XX PT Novel polynucleotide promoter sequences from Pine and Eucalyptus useful
PT for modifying expression of endogenous and/or heterologous
PT polynucleotides in transgenic plants

XX PS Claim 1; Page 68-69; 121pp; English.

XX CC The invention relates to isolated promoter sequences from Pinus radiata
CC and Eucalyptus grandis, comprising a sequence chosen from leaf-, root-,
CC flower-, pollen-, bud-, meristem-specific promoters or temporally
CC regulated promoters such as xylogenesis-specific promoters. The promoter
CC polypeptides and their related polynucleotides are useful in the
CC production of genetic constructs, used for modifying gene expression in a
CC target organism, in particular a plant. The method is useful for
CC modifying expression of a polynucleotide that comprises an intron
CC sequence, through removal of the intron sequence. The method is useful
CC for modifying growth and development of plants, and cellular responses to
CC external stimulus, such as environmental factors and disease pathogens.
CC The sequences are useful in genome and physical mapping, in positional
CC cloning of genes, in various assays to determine biological activity, to
CC raise antibodies, to isolate corresponding interacting proteins and other
CC compounds, and to quantitatively determine levels of interacting proteins
CC or other compounds. Sequences ABK17016-ABK17125 represent Pinus radiata
CC and Eucalyptus grandis polynucleotides and PCR primers used in the method
CC of the invention.

XX SQ Sequence 411 BP; 112 A; 118 C; 80 G; 101 T; 0 other;

Query Match 3.9%; Score 40.2; DB 24; Length 411;
Best Local Similarity 93.3%; Pred. NO. 0.029;
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ACTATAGGACCGGTGGTCGACGGCCGGGCTGGTATGAAGGTG 48
|||||

Db 1 ACTATAGGACCGGTGGTCGACGGCCGGGCTGGTATGAAGGTG 45
|||||

RESULT 12

AAX35652

ID AAX35652 standard; cDNA; 594 BP.

XX AC AAX35652;

XX DT 09-JUL-1999 (first entry)

XX DE 5' region of human heparanase cDNA.

XX KW Heparanase; hpa; modulator; heparin-binding growth factor;
KW cellular response; cytokine; cell interaction; plasma lipoprotein;
KW cellular susceptibility; infection; disintegration;
KW neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease; neutralise;
KW plasma heparin; micrometastasis; autoimmune lesion; renal failure;
KW ss.

XX OS Homo sapiens.

XX PN WO9911798-A1.

XX PD 11-MAR-1999.

XX PF 31-AUG-1998; 98WO-US17954.

XX PR 02-JUL-1998; 98US-0109386.

XX PR 02-SEP-1997; 97US-0922170.

XX PA (FRIE/) FRIEDMAN M M.

XX PA (HADA-) HADASIT MEDICAL RES SERVICES & DEV.

XX PA (INSI-) INSIGHT STRATEGY & MARKETING LTD.

XX PI Feinstein E, Pecker I, Vlodavsky I;

XX DR WPI; 1999-302255/25.

XX PT New human polynucleotide useful for treating angiogenesis,
PT restenosis, and inflammation

XX PS Example 8; Page 69; 63pp; English.

XX CC The specification describes a polypeptide having heparanase (hpa)
CC activity. The recombinant protein is used as a modulator of
CC heparin-binding growth factors, cellular responses to heparin-binding
CC growth factors and cytokines, cell interaction with plasma lipoproteins,
CC cellular susceptibility to viral, protozoal and bacterial infections
CC or disintegration of neurodegenerative plaques. Heparanase may be
CC useful for conditions such as wound healing, angiogenesis, restenosis,
CC atherosclerosis, inflammation, neurodegenerative diseases, and viral
CC infections. Mammalian heparanase can be used to neutralize plasma
CC heparin, and anti-heparanase antibodies may be applied for
CC immunodetection and diagnosis of micrometastases, autoimmune lesions,
CC and renal failure in biopsy specimens, plasma samples, and body fluids.
CC The present sequence represents the 5' sequence of human heparanase
CC cDNA.

XX SQ Sequence 594 BP; 133 A; 134 C; 209 G; 118 T; 0 other;

Query Match 3.9%; Score 40; DB 20; Length 594;
Best Local Similarity 100.0%; Pred. NO. 0.041;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTACTATAGGACCGGTGGTCGACGGCCGGGCTGGTAT 41
|||||

Db 2 TTACTATAGGACCGGTGGTCGACGGCCGGGCTGGTAT 41
|||||

RESULT 13

AAA75054

ID AAA75054 standard; cDNA; 594 BP.

XX AC AAA75054;

RESULT 12

```
DT 15-JAN-2001 (first entry)
XX Upstream sequence of human cDNA encoding heparanase.
DE
XX
XX Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; ds.
XX
XX Homo sapiens.
XX
XX WO200052178-A1.
XX
XX 08-SEP-2000.
XX
XX 14-FEB-2000; 2000WO-US03542.
XX
XX 01-MAR-1999; 99US-0258892.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX (FRIE/) FRIEDMAN M M.
XX
XX Pecker I, Vlodavsky I, Feinstein E;
XX
XX WPI; 2000-579289/54.
XX
XX New polynucleotides encoding a polypeptide having heparanase activity,
XX useful in wound healing and in gene therapy, particularly in treating
XX tumour, inflammation, autoimmunity, neurodegenerative diseases -
XX
XX Example 7; Page 127; 152pp; English.
XX
XX The present sequence is an upstream sequence of cDNA encoding encodes a
XX human protein with heparanase catalytic activity. The heparanase (hpa)
XX polynucleotide is useful in gene therapy, particularly in treating
XX tumour, inflammation or autoimmunity. Particularly, the polynucleotide
XX is useful in modulating the bioavailability of heparin-binding growth
XX factors, cellular responses to heparin-binding growth factors (e.g. bFGF)
XX and cytokines (e.g. interleukin (IL)-8), cell interaction with plasma
XX lipoproteins, cellular susceptibility to certain viral and some bacterial
XX and protozoa infections, or disintegration of neurodegenerative plaques.
XX The polynucleotide is also useful in wound healing (e.g. thermal,
XX chemical or radiation burns), and in the treatment of angiogenesis,
XX restenosis, atherosclerosis, inflammation, neurodegenerative diseases
XX (Gerstmann-Straussler Syndrome or Creutzfeldt-Jakob disease), and some
XX viral, bacterial or protozoa infections.
XX
XX Sequence 594 BP; 133 A; 134 C; 209 G; 118 T; 0 other;
SQ
Query Match 3.9%; Score 40; DB 21; Length 594;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TTACTATAGGCGCGCTGTCGACGGCCGGCTGGTAT 41
DB 2 TTACTATAGGCGCGCTGTCGACGGCCGGCTGGTAT 41
RESULT 14
AAV08328
ID AAV08328 standard; cDNA; 1680 BP.
XX
XX AAV08328;
XX
XX 04-FEB-1999 (first entry)
XX
XX Hyaluronate synthase promoter.
XX
XX Hyaluronate synthase; promoter; human;
XX hyaluronic acid production inhibitor; ds.
XX
```

```
OS Homo sapiens.
XX
XX JPI0295383-A.
XX
XX 10-NOV-1998.
XX
XX 24-APR-1997; 97JP-0107624.
XX
XX 24-APR-1997; 97JP-0107624.
XX
XX (ADSK-) ADVANCED SKIN RES KENKUSHO KK.
XX
XX WPI; 1999-038280/04.
XX
XX A promoter for hyaluronate synthase gene - used for screening of
XX modulators of the enzyme
XX
XX Claim 1; Page 5-6; 7pp; Japanese.
XX
XX This sequence represents the DNA of the invention, and can act as a
XX promoter for the hyaluronate synthase gene. It was isolated from human
XX cDNA sources. The DNA is useful for screening a drug enhancing or
XX inhibiting production of hyaluronic acid.
XX
XX Sequence 1680 BP; 452 A; 381 C; 467 G; 380 T; 0 other;
SQ
Query Match 3.9%; Score 40; DB 20; Length 1680;
Best Local Similarity 100.0%; Pred. No. 0.077;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTTACTATAGGCGCGCTGTCGACGGCCGGCTGGTA 40
DB 9 CTTACTATAGGCGCGCTGTCGACGGCCGGCTGGTA 48
RESULT 15
AAZ51547
ID AAZ51547 standard; DNA; 1224 BP.
XX
XX AAZ51547;
XX
XX 21-JUN-2000 (first entry)
XX
XX Seed-preferred promoter-2 from maize end gene.
XX
XX Seed-preferred promoter; maize; end gene; endosperm; endl; endl2;
XX fatty acid; starch profile; carbohydrate profile; amino acid content;
XX agronomic; insect resistance; disease resistance; herbicide resistance;
XX grain characteristic; ds.
XX
XX Zea mays.
XX
XX WO200012733-A1.
XX
XX 09-MAR-2000.
XX
XX 25-AUG-1999; 99WO-US19604.
XX
XX 28-AUG-1998; 98US-0098230.
XX
XX (PION-) PIONEER HI-BRED INT INC.
XX
XX Martino-Catt SJ, Lappegard KK, Olsen O, Linnestad C, Abbitt SE;
XX
XX WPI; 2000-256648/22.
XX
XX Maize promoter driving transcription in a seed-preferred manner, for
XX stably transforming plant cells -
XX
XX Claim 1; Pages 35-36; 43pp; English.
XX
XX The present sequence is a seed-preferred promoter isolated from maize
XX endosperm specific genes, endl or endl2. The promoter is capable of
```

CC driving transcription in a seed-preferred manner and can be used in an
CC expression cassette, to stably transform plant cells. The expression
CC cassette can be used to modify the fatty acid content of seeds, alter
CC the starch or carbohydrate profile, and/or alter the amino acid content
CC of the seed. It can also be used to deliver genes encoding important
CC traits for agronomics, including insect resistance, disease resistance,
CC herbicide resistance, and grain characteristics.

XX
SQ Sequence 1224 BP; 378 A; 239 C; 182 G; 425 T; 0 other;

Query Match 3.9%; Score 39.8; DB 21; Length 1224;
Best Local Similarity 95.3%; Pred. No. 0.074;
Matches 41; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGGTATGAAG 45

|||||
Db 1 TACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGGTAAAAG 43
|||||

Search completed: December 27, 2002, 22:42:34
Job time : 256 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 28, 2002, 00:01:58 ; Search time 54 Seconds
(without alignments)
5849.575 Million cell updates/sec

Title: US-09-945-376-3

Perfect score: 1030
Sequence: 1 ctctactataggcagcgctg.....tgtatatatacaagggtctgg 1030

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA: *

- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
- 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
- 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
- 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
- 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
- 6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42.4	4.1	6078	4	US-09-173-914-1
2	39.8	3.9	3718	4	US-09-424-283-6
3	39.6	3.8	7218	1	US-08-232-463-14
4	39	3.8	341	4	US-09-323-195A-1
5	39	3.8	4526	4	US-09-424-283-7
6	38.4	3.7	921	4	US-09-377-648-4
7	38.4	3.7	2791	4	US-09-570-367C-1
8	38	3.7	336	4	US-09-276-599-13
9	38	3.7	763	4	US-09-276-599-14
10	37.8	3.7	565	4	US-09-323-195A-5
11	37.8	3.7	1924	4	US-09-424-283-5
12	37.4	3.6	48	4	US-08-913-014A-18
13	37.4	3.6	2791	4	US-09-570-367C-1
14	37.2	3.6	1478	4	US-09-545-814-28
15	37.2	3.6	1478	4	US-09-545-814-30
16	37.2	3.6	1677	4	US-09-545-814-13
17	37.2	3.6	1677	4	US-09-545-814-15
18	37.2	3.6	1749	4	US-09-545-814-4
19	37.2	3.6	1749	4	US-09-545-814-6
20	37.2	3.6	1919	4	US-09-545-814-31
21	37.2	3.6	1919	4	US-09-545-814-33
22	37.2	3.6	2610	4	US-09-545-814-1
23	37.2	3.6	2610	4	US-09-545-814-3
24	37	3.6	340	4	US-09-323-195A-3
25	36	3.5	255	4	US-09-257-583-5
26	36	3.5	342	4	US-09-323-195A-6
27	36	3.5	555	4	US-08-905-223-37

28	36	3.5	555	4	US-09-247-155-37	Sequence 37, Appl
29	35	3.4	189	4	US-09-134-001C-1688	Sequence 1688, Ap
30	34.8	3.4	47	2	US-08-582-562A-53	Sequence 53, Appl
31	34.8	3.4	47	2	US-08-778-494B-53	Sequence 53, Appl
32	34.8	3.4	1881	2	US-08-596-300A-4	Sequence 4, Appl
33	34.8	3.4	3041	2	US-08-596-300A-5	Sequence 5, Appl
34	34.8	3.4	3718	2	US-08-596-300A-6	Sequence 6, Appl
35	34.8	3.4	3796	2	US-08-596-300A-3	Sequence 3, Appl
36	33	3.2	1440	4	US-09-134-001C-1291	Sequence 1291, Ap
37	32.2	3.1	340	4	US-09-323-195A-4	Sequence 4, Appl
38	32.2	3.1	396	4	US-08-887-534A-21	Sequence 21, Appl
39	32	3.1	500	3	US-08-755-587-37	Sequence 37, Appl
40	31.8	3.1	1218	3	US-08-976-122-1	Sequence 1, Appl
41	31.4	3.0	3064	4	US-09-600-776-5	Sequence 5, Appl
42	31.4	3.0	246240	2	US-08-724-394A-20	Sequence 20, Appl
43	31.4	3.0	246240	2	US-08-724-394A-21	Sequence 21, Appl
44	31.4	3.0	246240	2	US-08-724-394A-22	Sequence 22, Appl
45	31.2	3.0	6202	1	US-08-484-101B-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1
US-09-173-914-1
; Sequence 1, Application US/09173914
; Patent No. 6171857
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Eric
; TITLE OF INVENTION: A No. 6171857el Leucine zipper, KARP-1 and
; FILE OF INVENTION: Methods of Regulating DNA Dependent Protein Kinase Activity
; FILE REFERENCE: B0877/7017/HK
; CURRENT APPLICATION NUMBER: US/09/173,914
; CURRENT FILING DATE: 1998-10-16
; EARLIER APPLICATION NUMBER: 60/064,557
; EARLIER FILING DATE: 1997-10-17
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 6078
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (178)...(178)
; NAME/KEY: unsure
; LOCATION: (230)...(230)
; NAME/KEY: unsure
; LOCATION: (232)...(232)
; NAME/KEY: unsure
; LOCATION: (234)...(234)
; NAME/KEY: unsure
; LOCATION: (453)...(453)
; NAME/KEY: unsure
; LOCATION: (473)...(473)
; NAME/KEY: unsure
; LOCATION: (610)...(610)
; NAME/KEY: unsure
; LOCATION: (612)...(612)
; NAME/KEY: unsure
; LOCATION: (2175)...(2175)
; NAME/KEY: unsure
; LOCATION: (1014)...(1014)
US-09-173-914-1

Query Match 4.1%; Score 42.4; DB 4; Length 6078;

Best Local Similarity 97.7%; Pred. No. 0.0041;

Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCACGCGTGTGTCGACGGCCCGGGCTGGTATGAA 44

|||||

Db 9 CTTACTATAGGCACGCGTGTGTCGACGGCCCGGGCTGGTCTGAA 52

; PRIOR APPLICATION NUMBER: PCT/US98/10465
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: US 60/047,568
; PRIOR FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 4526
; TYPE: DNA
; ORGANISM: Glycine max
US-09-424-283-7

Query Match 3.8%; Score 39; DB 4; Length 4526;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTACTATAGGCACGGGTGTCGACGGCCGGCTGGTA 40
|||||
Db 97 TTACTATAGGCACGGGTGTCGACGGCCGGCTGGTA 135

RESULT 6

US-09-377-648-4/c
; Sequence 4, Application US/09377648
; Patent No. 6225529
; GENERAL INFORMATION:
; APPLICANT: Lappegard, Kathryn
; APPLICANT: Martino-Catt, Susan
; TITLE OF INVENTION: Seed-Preferred Promoters
; FILE REFERENCE: 0869
; CURRENT APPLICATION NUMBER: US/09/377,648
; CURRENT FILING DATE: 1999-08-19
; EARLIER APPLICATION NUMBER: US 60/097,233
; EARLIER FILING DATE: 1998-08-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)...(922)
US-09-377-648-4

Query Match 3.7%; Score 38.4; DB 4; Length 921;
Best Local Similarity 97.5%; Pred. No. 0.028;
Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCACGGGTGTCGACGGCCGGCTGGTA 40
|||
Db 912 CTCACTATAGGCACGGGTGTCGACGGCCGGCTGGTA 873

RESULT 7

US-09-570-367C-1
; Sequence 1, Application US/09570367C
; Patent No. 6338851
; GENERAL INFORMATION:
; APPLICANT: Gorczynski, Reginald M.
; TITLE OF INVENTION: Methods and Compositions for Immunomodulation
; FILE REFERENCE: 9579-21
; CURRENT APPLICATION NUMBER: US/09/570,367C
; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/064,764
; PRIOR FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2791
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-570-367C-1

Query Match 3.7%; Score 38.4; DB 4; Length 2791;
Best Local Similarity 87.5%; Pred. No. 0.051;
Matches 42; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 4 ACTATAGGCACGGGTGTCGACGGCCGGCTGGTATGAAGTGGGA 51
|||||
Db 1 ACTATAGGCACGGGTGTCGACGGCCGGCTGGTACTGAGAAGAA 48

RESULT 8

US-09-276-599-13/c
; Sequence 13, Application US/09276599
; Patent No. 6380459
; GENERAL INFORMATION:
; APPLICANT: Perera, J. Ranjan
; APPLICANT: Rice, Stephen J.
; TITLE OF INVENTION: Composition and methods for the
; FILE REFERENCE: 11000.1036
; CURRENT APPLICATION NUMBER: US/09/276,599
; CURRENT FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-276-599-13

Query Match 3.7%; Score 38; DB 4; Length 336;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTACTATAGGCACGGGTGTCGACGGCCGGCTGGT 39
|||||
Db 334 TTACTATAGGCACGGGTGTCGACGGCCGGCTGGT 297

RESULT 9

US-09-276-599-14/c
; Sequence 14, Application US/09276599
; Patent No. 6380459
; GENERAL INFORMATION:
; APPLICANT: Perera, J. Ranjan
; APPLICANT: Rice, Stephen J.
; TITLE OF INVENTION: Composition and methods for the
; FILE REFERENCE: 11000.1036
; CURRENT APPLICATION NUMBER: US/09/276,599
; CURRENT FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 763
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-276-599-14

Query Match 3.7%; Score 38; DB 4; Length 763;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTACTATAGGCACGGGTGTCGACGGCCGGCTGGT 39
|||||
Db 761 TTACTATAGGCACGGGTGTCGACGGCCGGCTGGT 724

RESULT 10

US-09-323-195A-5
; Sequence 5, Application US/09323195A
; Patent No. 6462257
; GENERAL INFORMATION:

```
; APPLICANT: Pullman, Gerald
; APPLICANT: Cairney, John
; APPLICANT: Perrera, Rajan
; TITLE OF INVENTION: VICILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND
; TITLE OF INVENTION: METHODS OF USING THE SAME
; FILE REFERENCE: IPST0009
; CURRENT APPLICATION NUMBER: US/09/323,195A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 565
; TYPE: DNA
; ORGANISM: Pinus taeda
; FEATURE:
; OTHER INFORMATION: n at 489 is a, C, g, or t
; OTHER INFORMATION: n at 503 is a, C, g, or t
; OTHER INFORMATION: n at 504 is a, C, g, or t
; OTHER INFORMATION: n at 522 is a, C, g, or t
; OTHER INFORMATION: n at 533 is a, C, g, or t
; OTHER INFORMATION: n at 543 is a, C, g, or t
; OTHER INFORMATION: n at 549 is a, C, g, or t
; OTHER INFORMATION: n at 564 is a, C, g, or t
US-09-323-195A-5

Query Match          3.7%; Score 37.8; DB 4; Length 565;
Best Local Similarity 95.1%; Pred. No. 0.033;
Matches 39; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ACTATAGGCACCGCTGGTGCAGCGCCGGCGGTGGTATGAA 44
|||
Db 1 ACTATAGGCACCGCTGGTGCAGCGCCGGCGGTGGTAAAAA 41
|||

RESULT 11
US-09-424-283-5/c
; Sequence 5, Application US/09424283
; Patent No. 6437219
; GENERAL INFORMATION:
; APPLICANT: Grimes, et al.
; TITLE OF INVENTION: Sucrose binding proteins
; FILE REFERENCE: 4630-50206
; CURRENT APPLICATION NUMBER: US/09/424,283
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: PCT/US98/10465
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: US 60/047,568
; PRIOR FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1924
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (119)..(1588)
US-09-424-283-5

Query Match          3.7%; Score 37.8; DB 4; Length 1924;
Best Local Similarity 95.1%; Pred. No. 0.065;
Matches 39; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCACCGCTGGTGCAGCGCCGGCGGTGGTAT 41
|||
Db 1803 CTCACATAGGCACCGCTGGTGCAGCGCCGGCGGTGGT 1763
|||

RESULT 12
US-08-913-014A-18
; Sequence 18, Application US/08913014A
; Patent No. 6235878
; GENERAL INFORMATION:
```

```
; APPLICANT: Nishi, Kazunori
; APPLICANT: Hikichi, Yukiko
; APPLICANT: Shincani, Yasushi
; TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
; TITLE OF INVENTION: PRODUCTION AND USE
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David G. Conlin, Esq.
; ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
; STREET: 130 Water Street
; CITY: Boston,
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/913,014A
; FILING DATE: 04-SEP-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP97/02480
; FILING DATE: July 17, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: David G. Conlin
; REGISTRATION NUMBER: 27,026
; REFERENCE/DOCKET NUMBER: 342/47694
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-523-3400
; TELEFAX: 617-523-6440
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48
; TYPE: Nucleic acid
; STRANDEDNESS: Single
; TOPOLOGY: Linear
; MOLECULE TYPE: Synthetic DNA
US-08-913-014A-18

Query Match          3.6%; Score 37.4; DB 4; Length 48;
Best Local Similarity 97.4%; Pred. No. 0.012;
Matches 38; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCACCGCTGGTGCAGCGCCGGCGGTGGT 39
|||
Db 10 CTCACTATAGGCACCGCTGGTGCAGCGCCGGCGGTGGT 48
|||

RESULT 13
US-09-570-367C-1/c
; Sequence 1, Application US/09570367C
; Patent No. 6338851
; GENERAL INFORMATION:
; APPLICANT: Gorczynski, Reginald M.
; TITLE OF INVENTION: Methods and Compositions for Immunomodulation
; FILE REFERENCE: 9579-21
; CURRENT APPLICATION NUMBER: US/09/570,367C
; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/064,764
; PRIOR FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2791
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-570-367C-1

Query Match          3.6%; Score 37.4; DB 4; Length 2791;
```

Best Local Similarity 97.4%; Pred. No. 0.11;
Matches 38; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ACTATAGGACGGCTGGTCTGACGCGCGGGCTGATG 42
Db 2791 ACTATAGGACGGCTGGTCTGACGCGCGGGCTGATG 2753

RESULT 14
US-09-545-814-28/c
; Sequence 28, Application US/09545814
; Patent No. 6416977
; GENERAL INFORMATION:
; APPLICANT: Becher, Anna M.
; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
; FILE REFERENCE: FC-5-C1
; CURRENT APPLICATION NUMBER: US/09/545,814
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,833
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 1478
; TYPE: DNA
; ORGANISM: Tagged Ctenocephalides felis
; NAME/KEY: CDS
; LOCATION: (1)..(1467)
US-09-545-814-28

Query Match 3.6%; Score 37.2; DB 4; Length 1478;
Best Local Similarity 47.1%; Pred. No. 0.087;
Matches 114; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 66 ATATAGCTGTGAGAGATACATCAATTCACAAGCTTCGATGTCACATACAGCCTACAGC 125
Db 1023 AGACATTACGTGGTAGCATCTAATAATTCGCAAGCTTCTGTACATGATATCCTTCTCTG 1082

QY 126 CATTCGAGCAGGCGCTAGGACGCGCTCAATAACTTGGAGGAATGCCAAGAATGTG 185
Db 1083 CAGCCTAAATTTGGCAATCGGACAGCCATGGTGATCTCCAGTTTTTGGCTTTGTTGT 1142

QY 186 GATTACAAACAGTTCTATCTGAGCGACAGCTCTAAACGAGCTTTGTATAGTCAATTTTATC 245
Db 1143 GAAAGCCCTTCGCAATTTCTTGGACAAAGTATAAAATTTCTTTGTCGGAATACTGCC 1202

QY 246 TCCCAACCGGCATTAGTAAGCAGGCGCCCAACGCAATTCAAATTCGATGTAATGTCTAC 305
Db 1203 ACCTCTATCAGCAGCAGGACTACTCCCAATCCAGTCGAAACCGTCGAACCTGTATTTC 1262

QY 306 TT 307
Db 1263 GT 1264

Search completed: December 28, 2002, 01:26:51
Job time : 61 secs

Best Local Similarity 97.4%; Pred. No. 0.11;
Matches 38; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ACTATAGGACGGCTGGTCTGACGCGCGGGCTGATG 42
Db 2791 ACTATAGGACGGCTGGTCTGACGCGCGGGCTGATG 2753

RESULT 14
US-09-545-814-28/c
; Sequence 28, Application US/09545814
; Patent No. 6416977
; GENERAL INFORMATION:
; APPLICANT: Becher, Anna M.
; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
; FILE REFERENCE: FC-5-C1
; CURRENT APPLICATION NUMBER: US/09/545,814
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,833
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 1478
; TYPE: DNA
; ORGANISM: Tagged Ctenocephalides felis
; NAME/KEY: CDS
; LOCATION: (1)..(1467)
US-09-545-814-28

Query Match 3.6%; Score 37.2; DB 4; Length 1478;
Best Local Similarity 47.1%; Pred. No. 0.087;
Matches 114; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 66 ATATAGCTGTGAGAGATACATCAATTCACAAGCTTCGATGTCACATACAGCCTACAGC 125
Db 456 AGACATTAGTGGTAGCATCTAATAATTCGCAAGCTTCTGTACATGATATCCTTCTCTG 397

QY 126 CATTCGAGCAGGCGCTAGGACGCGCTCAATAACTTGGAGGAATGCCAAGAATGTG 185
Db 396 CAGCCTAAATTTGGCAATCGGACAGCCATGGTGATCTCCAGTTTTTGGCTTTGTTGT 337

QY 186 GATTACAAACAGTTCTATCTGAGCGAGCTTAACGCAAGCTTTGTATAGTCAATTTTATC 245
Db 336 GAAAGCCCTTCGCAATTTCTTGGACAAAGTATAAAATTTGTCTTTGTCGGAATACTGCC 277

QY 246 TCCCAACCGGCATTAGTAAGCAGGCGCCCAACGCAATTCAAATTCGATGTAATGTCTAC 305
Db 276 ACCTCTATCAGCAGCAGGACTACTCCCAATCCAGTCGAAACCGTCGAACCTGTATTTC 217

QY 306 TT 307
Db 216 GT 215

RESULT 15
US-09-545-814-30
; Sequence 30, Application US/09545814
; Patent No. 6416977
; GENERAL INFORMATION:
; APPLICANT: Becher, Anna M.
; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
; FILE REFERENCE: FC-5-C1
; CURRENT APPLICATION NUMBER: US/09/545,814
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,833
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2002, 23:38:28 ; Search time 252 Seconds
(without alignments)
9204.596 Million cell updates/sec

Title: US-09-945-376-3
Perfect score: 1030
Sequence: 1 ctctattaggccgcgcgtg.....tgatatatacaaggtctgg 1030

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	45	4.4	3548	AAD38882	Alfalfa plastocyan
2	42.4	4.1	6078	AAF31861	Human KARP-1 DNA
3	41	4.0	2614	AAV40621	Cotton promoter re
4	40.6	3.9	1038	AAC62796	Senescence-like pr
5	40.6	3.9	1038	ABK17061	Pinus radiata prom
6	40.4	3.9	598	AAK15016	Rice anther-specif
7	40.4	3.9	2096	AAC62810	O-methyl transfera
8	40.4	3.9	2096	ABK17075	Eucalyptus grandis
C 9	40.4	3.9	2611	ABA91255	Arabidopsis BONSAI

10	40.2	3.9	411	21	AAC62780	Flower specific pr
11	40.2	3.9	411	24	ABK17045	Eucalyptus grandis
12	40	3.9	594	20	AAK35652	5' region of human
13	40	3.9	594	21	AAA75054	Upstream sequence
14	40	3.9	1680	20	AAV08328	Hyaluronate synth
15	39.8	3.9	1224	21	AAZ51547	Seed-preferred pro
16	39.8	3.9	2385	21	AAC83331	PART-1 promoter re
17	39.8	3.9	2475	24	ABL50287	Rat glyceraldehyde
18	39.8	3.9	2565	24	ABL50285	Rat glyceraldehyde
19	39.8	3.9	3718	20	AAV72244	G. max SBP2 DNA 5'
20	39.8	3.9	3775	24	AAD38881	Alfalfa AP2 adapto
21	39.4	3.8	288	21	AAC62788	Isoflavone reducta
22	39.4	3.8	288	24	ABK17053	Eucalyptus grandis
23	39.4	3.8	515	21	AAC62816	MIF homologue codi
24	39.4	3.8	515	24	ABK17081	Eucalyptus grandis
25	39.4	3.8	2492	22	AAF61948	T. thermophila del
26	39.4	3.8	2867	19	AAV68599	Flower specific pr
27	39	3.8	311	21	AAC62809	Lecithin-cholester
28	39	3.8	311	24	ABK17074	Eucalyptus grandis
29	39	3.8	516	24	AAK15015	Rice tubulin-like
30	39	3.8	927	21	AAC62779	Flower specific pr
31	39	3.8	927	24	ABK17044	Eucalyptus grandis
32	39	3.8	940	24	AAD38884	Alfalfa plastocyan
33	39	3.8	1126	21	AAC62802	Pollen specific pr
34	39	3.8	1126	24	ABK17067	Pinus radiata prom
35	39	3.8	1241	19	AAV18017	Pinus radiata cone
C 36	39	3.8	1242	18	AAV94308	PfP1 gene promote
37	39	3.8	4526	20	AAV72245	G. max SBP1 DNA 5'
38	38.6	3.7	1196	22	AAI68306	Aspergillus oryzae
C 39	38.6	3.7	5031	22	AAF76068	Maize MADS-box gen
C 40	38.6	3.7	5031	22	AAF76455	Maize ZmMADS2 gene
C 41	38.4	3.7	382	21	AAC62789	Glyceraldehyde-3-p
C 42	38.4	3.7	382	21	ABK17054	Eucalyptus grandis
C 43	38.4	3.7	648	21	AAC62787	Isoflavone reducta
C 44	38.4	3.7	648	24	ABK17052	Eucalyptus grandis
C 45	38.4	3.7	921	21	AAZ51580	Maize seed-prefer

ALIGNMENTS

RESULT 1
AAD38882/c
ID AAD38882 standard; DNA; 3548 BP.
XX AAD38882;
AC AAD38882;
XX 23-SEP-2002 (first entry)
DT Alfalfa plastocyanin gene coding fragment.
DE Alfalfa plastocyanin gene coding fragment.
XX Expression regulatory sequence; recombinant polypeptide; gene cloning;
KW alfalfa; plastocyanin; ds.
XX Medicago sativa.
XX WO200236786-A2.
XX 10-MAY-2002.
XX 31-OCT-2001; 2001WO-CA01532.
XX 31-OCT-2000; 2000US-244214P.
XX (MEDI-) MEDICAGO INC.
XX Vezina L, D'aoust M, Arcand F, Bilodeau P;
XX WPI; 2002-471503/50.
XX Isolating and characterizing an expression regulatory sequence for
XX expressing recombinant polypeptides and/or RNA, comprises producing
XX oligonucleotide primers that amplify sequences upstream or downstream

PT of cDNAs -
 PS Example 2; Page 68-69; 74pp; English.
 XX
 CC The invention relates to a method for isolating and characterising an
 CC expression regulatory sequence for the expression of recombinant
 CC polypeptides and/or RNA. The method comprising producing at least one
 CC oligonucleotide primer from cDNAs of a cDNA library, where the
 CC oligonucleotide primer allows amplification of genomic sequences upstream
 CC or downstream of the cDNAs. The method is useful for isolating,
 CC characterising and identifying a large number of known and unknown
 CC promoters that are active under any desired environmental condition to
 CC which a cell may be exposed and not just to the exemplified isolation of
 CC promoters that are capable of expression in specific conditions. The
 CC methods are also useful for cloning genes from any host, or from a
 CC specific tissue with such host, from which a cDNA library may be
 CC constructed; for the identification and isolation of analogous promoters,
 CC signal peptides and structural genes in several species of multicellular
 CC and unicellular organisms and as a high throughput identification system
 CC of candidate therapeutic targets. The promoter sequences may be used to
 CC regulate the synthesis of polypeptides. The present sequence is alfalfa
 CC plastocyanin gene coding fragment.
 XX
 SQ Sequence 3548 BP; 1112 A; 576 C; 727 G; 1133 T; 0 other;
 Query Match 4.4%; Score 45; DB 24; Length 3548;
 Best Local Similarity 90.6%; Pred. No. 0.0029;
 Matches 48; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 TTACTATAGGCACGGTGGTCGACGGCCCGGCTGGTATGAAGTGGGAACC 54
 Db 3529 TTACTATAGGCACGGTGGTCGACGGCCCGGCTGGTATGAAGTGGCAAC 3477
 RESULT 2
 AAF31861
 ID AAF31861 standard; DNA; 6078 BP.
 XX
 AC AAF31861;
 XX
 DT 12-APR-2001 (first entry)
 XX
 DE Human KARP-1 DNA.
 XX
 KW Human; KARP-1; Ku86 autoantigen related protein; cancer;
 KW immune deficiency disorder; biliary tract cancer; leucine zipper protein;
 KW cytostatic; immunosuppressant; gene therapy; KARP-1 inhibitor; ds.
 XX
 OS Homo sapiens.
 XX
 PN US6171857-B1.
 XX
 PD 09-JAN-2001.
 XX
 PF 16-OCT-1998; 98US-0173914.
 XX
 PR 17-OCT-1997; 97US-0064557.
 XX
 PA (UYBR-) UNIV BROWN RES FOUND.
 XX
 PI Hendrickson EA;
 XX
 DR WPI; 2001-146208/15.
 DR P-PSDB; AAB66590.
 XX
 PT Novel nucleic acids encoding leucine zipper protein, KARP-1
 PT polypeptide, useful for treating cancer and immune deficiency disorder
 PT
 XX Claim 1; Column 51-56; 61pp; English.
 XX
 CC The present sequence is given in a specification relating to an isolated
 CC Ku86 Autoantigen Related Protein (KARP-1) nucleic acid molecule. The

CC KARP-1 nucleic acid and KARP-1 protein are useful for the treatment
 CC and/or diagnosis of diseases such as cancer and immune deficiency
 CC disorders. They are useful in combination with a KARP-1 inhibitor that
 CC inhibits double stranded DNA base repair. Inhibitors of KARP-1 are
 CC useful in the diagnosis or treatment of conditions characterised
 CC by the loss of KARP-1 activity and in the treatment of cancer,
 CC e.g. biliary tract cancer.
 XX
 SQ Sequence 6078 BP; 1806 A; 1301 C; 1392 G; 1543 T; 36 other;
 Query Match 4.1%; Score 42.4; DB 22; Length 6078;
 Best Local Similarity 97.7%; Pred. No. 0.028;
 Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 CTTTACTATAGGCACGGTGGTCGACGGCCCGGCTGGTATGAA 44
 Db 9 CTTTACTATAGGCACGGTGGTCGACGGCCCGGCTGGTCTGAA 52
 RESULT 3
 AAV40621
 ID AAV40621 standard; DNA; 2614 BP.
 XX
 AC AAV40621;
 XX
 DT 26-OCT-1998 (first entry)
 XX
 DE Cotton promoter region from an expansin gene.
 XX
 KW Cotton fibre expansin promoter; transformation; transgenic; ss.
 XX
 OS Gossypium hirsutum
 XX
 PN WO9830698-A1.
 XX
 PD 16-JUL-1998.
 XX
 PF 07-JAN-1998; 98WO-US00151.
 XX
 PR 07-JAN-1997; 97US-0034914.
 XX
 PA (CALJ) CALGENE INC.
 XX
 PI Pear JR, Stalker DM;
 XX
 DR WPI; 1998-399143/34.
 XX
 PT New DNA containing the promoter of the cotton expansin gene - used
 PT for tissue-selective expression of genes that alter cotton fibre
 PT phenotype
 XX
 PS Claim 2; Fig 1a-d; 26pp; English.
 XX
 CC The present sequence represents an expansin promoter region from
 CC Gossypium hirsutum. This has been sequenced and found to be
 CC approximately 2200 bp in length, and is immediately 5' to the cotton
 CC fibre expansin coding region. The promoter can be used in cotton
 CC transformation by linking it to a gene of interest, thus enabling the
 CC modification of transgenic cotton fibre cells.
 XX
 SQ Sequence 2614 BP; 892 A; 436 C; 401 G; 882 T; 3 other;
 Query Match 4.0%; Score 41; DB 19; Length 2614;
 Best Local Similarity 100.0%; Pred. No. 0.048;
 Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CTTTACTATAGGCACGGTGGTCGACGGCCCGGCTGGTAT 41
 Db 83 CTTTACTATAGGCACGGTGGTCGACGGCCCGGCTGGTAT 123
 RESULT 4
 AAC62796


```
ID AAC62796 standard; DNA; 1038 BP.
XX
AC AAC62796;
XX
XT 02-FEB-2001 (first entry)
XX
DE Senescence-like protein promoter coding sequence #1.
XX
KW Promoter; eucalyptus; pine; gene transcription; ds.
XX
OS Pinus radiata.
XX
PN WO200058474-A1.
XX
PD 05-OCT-2000.
XX
PF 24-FEB-2000; 2000WO-NZ00018.
XX
PR 25-MAR-1999; 99US-0276599.
XX
PS 30-JUL-1999; 99US-0146591.
XX
PA (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS LTD.
XX
PI Perera R, Rice SJ, Eagleton CK;
XX
DR WPI: 2000-647236/62.
DR P-PSDB; AAB28136.
XX
PT Novel promoter sequences useful for modulating transcription of plant
PT DNA sequences of interest and production of polypeptides -
XX
PS Claim 1; Pages 57-58; 93pp; English.
XX
CC The present invention relates to promoter sequences from eucalyptus
CC and pine. The present sequence is one such promoter. This sequence is
CC useful for modulating the transcription of DNA sequences of interest. The
CC sequences may also be used to tag or identify an organism or its
CC reproductive material.
XX
SQ Sequence 1038 BP; 298 A; 220 C; 233 G; 287 T; 0 other;

Query Match 3.9%; Score 40.6; DB 21; Length 1038;
Best Local Similarity 71.1%; Pred. No. 0.037;
Matches 69; Conservative 0; Mismatches 24; Indels 4; Gaps 1;

QY 2 TTACTATAGGCGACCGCTGGTCGACGGCCGGCTGGTATGAAGGTGGGAACCTCACTGG 61
Db 4 TTACTATAGGCGACCGCTGGTCGACGGCCGGCTGGTATAA----TGAGAACATGATAAG 59

QY 62 ATGCATATACTGCTGAGAGATAACATCAACAATTCA 98
Db 60 CTGTGTAATTCATGCTAGTACACCAATAACTTTTCTCA 96

RESULT 5
ABK17061
ID ABK17061 standard; cDNA; 1038 BP.
XX
AC ABK17061;
XX
DT 26-MAR-2002 (first entry)
XX
DE Pinus radiata promoter polynucleotide #20.
XX
KW Promoter; pine; leaf; root; flower; pollen; bud; meristem; xylogenesis;
KW temporally regulated promoter; Pinus radiata; Eucalyptus grandis; ss;
KW PCR primer.
XX
OS Pinus radiata.
XX
PN WO200198485-A1.
XX
```

```
PD 27-DEC-2001.
XX
XX 20-JUN-2001; 2001WO-NZ00115.
XX
PF 20-JUN-2000; 2000US-0598401.
XX
PR 28-NOV-2000; 2000US-0724624.
XX
XX (GENE-) GENESIS RES & DEV CORP LTD.
PA (FLET-) FLETCHER CHALLENGE FORESTS IND LTD.
XX
PI Perera R, Rice S, Eagleton C, Lasham A;
XX
DR WPI: 2002-114583/15.
DR P-PSDB; AAU80754.
XX
XX Novel polynucleotide promoter sequences from pine and Eucalyptus useful
XX for modifying expression of endogenous and/or heterologous
XX polynucleotides in transgenic plants -
XX
PS Claim 1; Page 73-74; 121pp; English.
XX
CC The invention relates to isolated promoter sequences from Pinus radiata
CC and Eucalyptus grandis, comprising a sequence chosen from leaf-, root-,
CC flower-, pollen-, bud-, meristem-specific promoters or temporally
CC regulated promoters such as xylogenesis-specific promoters. The promoter
CC polypeptides and their related polynucleotides are useful in the
CC production of genetic constructs, used for modifying gene expression in a
CC target organism, in particular a plant. The method is useful for
CC modifying expression of a polynucleotide that comprises an intron
CC sequence, through removal of the intron sequence. The method is useful
CC for modifying growth and development of plants, and cellular responses to
CC external stimulus, such as environmental factors and disease pathogens.
CC The sequences are useful in genome and physical mapping, in positional
CC cloning of genes, in various assays to determine biological activity, to
CC raise antibodies, to isolate corresponding interacting proteins and other
CC compounds, and to quantitatively determine levels of interacting proteins
CC or other compounds. Sequences ABK17016-ABK17125 represent Pinus radiata
CC and Eucalyptus grandis polynucleotides and PCR primers used in the method
CC of the invention.
XX
SQ Sequence 1038 BP; 298 A; 220 C; 233 G; 287 T; 0 other;

Query Match 3.9%; Score 40.6; DB 24; Length 1038;
Best Local Similarity 71.1%; Pred. No. 0.037;
Matches 69; Conservative 0; Mismatches 24; Indels 4; Gaps 1;

QY 2 TTACTATAGGCGACCGCTGGTCGACGGCCGGCTGGTATGAAGGTGGGAACCTCACTGG 61
Db 4 TTACTATAGGCGACCGCTGGTCGACGGCCGGCTGGTATAA----TGAGAACATGATAAG 59

QY 62 ATGCATATACTGCTGAGAGATAACATCAACAATTCA 98
Db 60 CTGTGTAATTCATGCTAGTACACCAATAACTTTTCTCA 96

RESULT 6
AAS15016
ID AAS15016 standard; cDNA; 598 BP.
XX
AC AAS15016;
XX
DT 14-FEB-2002 (first entry)
XX
DE Rice anther-specific promoter #2.
XX
KW Rice; promoter; gene therapy; transcription; monocot; anther;
KW chromosome mapping; gene mapping; antisense technology;
KW plant genetic engineering; ss.
XX
OS Oryza sativa.
XX
PN WO200181606-A2.
XX
```

PD 01-NOV-2001.
 XX PF 26-APR-2001; 2001WO-US13544.
 XX PR 26-APR-2000; 2000US-199870P.
 PR 12-JUL-2000; 2000US-217891P.
 PR 13-JUL-2000; 2000US-218366P.
 PR 23-AUG-2000; 2000US-227231P.
 PR 03-OCT-2000; 2000US-237736P.
 PR 29-NOV-2000; 2000US-253925P.
 XX (AKKA-) AKKADIX CORP.
 PA Perera JR, Lu M, Ray A;
 XX WPI; 2002-041419/05.
 XX Rice promoter sequences (I) useful in plant genetic engineering and
 PT molecular biology studies -
 PT Claim 1; Fig 1B; 47pp; English.
 PS The invention relates to plant polynucleotide sequences from rice which
 CC encode promoter components of the cellular activation and transcription
 CC apparatus (5' cis regulatory DNA sequences). The sequences may be used in
 CC the modification of gene activation and/or expression in eukaryotes,
 CC especially rice and other monocots. They have many applications in
 CC standard molecular biology, e.g. as for chromosome and gene mapping, in
 CC polymerase chain reaction (PCR) technology, for the study of gene
 CC function and expression in vivo, specific down-regulation of target
 CC genes, and in the production of sense and/or antisense nucleic acids.
 CC The present sequence represents rice anther-specific promoter #2
 CC as described in the invention.
 XX Sequence 598 BP; 174 A; 131 C; 113 G; 179 T; 1 other;
 SQ
 Query Match 3.9%; Score 40.4; DB 24; Length 598;
 Best Local Similarity 75.8%; Pred. No. 0.031;
 Matches 50; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
 QY 4 ACTATAGGCGACGCTGTCGACGCCGCGCTGTATCAAGTGGGAACCTCACTGGAT 63
 Db |||||
 QY 1 ACTATAGGCGACGCTGTCGACGCCGCGCTGTATCAAACTGCTCAAAATGTCGAT 60
 Db |||||
 QY 64 GCATAT 69
 Db |||||
 QY 61 TAATAT 66
 Db |||||
 RESULT 7
 AAC62810
 ID AAC62810 standard; DNA; 2096 BP.
 AC AAC62810;
 XX
 DT 02-FEB-2001 (first entry)
 XX
 DE O-methyl transferase promoter coding sequence #2.
 XX Promoter; eucalyptus; pine; gene transcription; ds.
 KW Eucalyptus grandis.
 OS WO2000058474-A1.
 PN 05-OCT-2000.
 XX 24-FEB-2000; 2000WO-NZ00018.
 PF 25-MAR-1999; 99US-0276599.
 PR 30-JUL-1999; 99US-0146591.
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 PA
 (FLET-) FLETCHER CHALLENGE FORESTS LTD.
 PA Perera R, Rice SJ, Eagleton CK;
 XX WPI; 2000-647236/62.
 DR P-PSDB; AAB28142.
 XX Novel promoter sequences useful for modulating transcription of plant
 PT DNA sequences of interest and production of polypeptides -
 PT Claim 1; Pages 61-62; 93pp; English.
 PS The present invention relates to promoter sequences from eucalyptus
 CC and pine. The present sequence is one such promoter. This sequence is
 CC useful for modulating the transcription of DNA sequences of interest. The
 CC sequences may also be used to tag or identify an organism or its
 CC reproductive material.
 XX Sequenced 2096 BP; 537 A; 535 C; 539 G; 485 T; 0 other;
 SQ
 Query Match 3.9%; Score 40.4; DB 21; Length 2096;
 Best Local Similarity 97.6%; Pred. No. 0.065;
 Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 2 TTACTATAGGCGACGCTGTCGACGCCGCGCTGTATGA 43
 Db |||||
 QY 3 TTACTATAGGCGACGCTGTCGACGCCGCGCTGTATGA 44
 Db |||||
 RESULT 8
 ABK17075
 ID ABK17075 standard; cDNA; 2096 BP.
 XX
 AC ABK17075;
 XX
 DT 26-MAR-2002 (first entry)
 XX Eucalyptus grandis promoter polynucleotide #31.
 DE Promoter; pine; leaf; root; flower; pollen; bud; meristem; xylogenesis;
 KW temporally regulated promoter; Pinus radiata; Eucalyptus grandis; ss;
 KW PCR primer.
 OS Eucalyptus grandis.
 XX WO200138485-A1.
 PN 27-DEC-2001.
 PD 20-JUN-2001; 2001WO-NZ00115.
 PF 20-JUN-2000; 2000US-0598401.
 PR 28-NOV-2000; 2000US-0724624.
 XX (GENE-) GENESIS RES & DEV CORP LTD.
 PA (FLET-) FLETCHER CHALLENGE FORESTS IND LTD.
 XX Perera R, Rice S, Eagleton C, Lasham A;
 XX WPI; 2002-114583/15.
 DR P-PSDB; AAU80760.
 XX Novel polynucleotide promoter sequences from pine and Eucalyptus useful
 PT for modifying expression of endogenous and/or heterologous
 PT polynucleotides in transgenic plants -
 PT Claim 1; Page 78; 121pp; English.
 PS The invention relates to isolated promoter sequences from Pinus radiata
 CC and Eucalyptus grandis, comprising a sequence chosen from leaf-, root-,
 CC flower-, pollen-, bud-, meristem-specific promoters or temporally
 CC regulated promoters such as xylogenesis-specific promoters. The promoter
 CC polypeptides and their related polynucleotides are useful in the

CC production of genetic constructs, used for modifying gene expression in a
CC target organism, in particular a plant. The method is useful for
CC modifying expression of a polynucleotide that comprises an intron
CC sequence, through removal of the intron sequence. The method is useful
CC for modifying growth and development of plants, and cellular responses to
CC external stimulus, such as environmental factors and disease pathogens.
CC The sequences are useful in genome and physical mapping, in positional
CC cloning of genes, in various assays to determine biological activity, to
CC raise antibodies, to isolate corresponding interacting proteins and other
CC compounds, and to quantitatively determine levels of interacting proteins
CC or other compounds. Sequences ABK17016-ABK17125 represent Pinus radiata
CC and Eucalyptus grandis polynucleotides and PCR primers used in the method
CC of the invention.

XX
SQ Sequence 2096 BP; 537 A; 535 C; 539 G; 485 T; 0 other;

Query Match 3.9%; Score 40.4; DB 24; Length 2096;
Best Local Similarity 97.6%; Pred. No. 0.065;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTACTATAGGCGACGGTGTGTCGACGGCCCGGGCTGGTATGA 43

Db 3 TTACTATAGGCGACGGTGTGTCGACGGCCCGGGCTGGTCTGA 44

RESULT 9

ABA91255/c

ID ABA91255 standard; DNA; 2611 BP.

AC ABA91255;

XX 04-APR-2002 (first entry)

XX Arabidopsis BONSAI1 (BON1) gene, involved in growth homeostasis.

XX BONSAI1: BON1; phospholipid binding protein; growth; homeostasis;
XX thermotolerance; transgenic plant; plant; gene; ds.

XX Arabidopsis thaliana.

XX WO200200697-A2.

XX 03-JAN-2002.

XX 25-JUN-2001; 2001WO-0520172.

XX 23-JUN-2000; 2000US-213863P.

XX (WHED) WHITEHEAD INST BIOMEDICAL RES.

XX Hua J, Grisafi P, Fink GR;

XX WPI; 2002-139899/18.

XX New phospholipid binding proteins and nucleic acids, useful for
XX modulating plant growth homeostasis, controlling cell expansion and
XX cell division, or producing plants where larger fruits and increased
XX biomass are desired

XX Disclosure; Fig 8A; 78pp; English.

XX The present sequence is that of the BONSAI1 (BON1) gene of
XX Arabidopsis thaliana. The coding region of the BON1 gene is
XX given in ABA91256. The BON1 gene permits wild-type Arabidopsis
XX plants to maintain a relatively constant size over a wide range of
XX temperatures. Thus, bon1 null mutants produce miniature fertile
XX plants at 22 degrees C, but a wild-type phenotype at 28 degrees C.
XX BON1 has a direct role in regulating cell expansion and cell
XX division at temperatures lower than those at which Arabidopsis is
XX normally grown. The BON1 protein (see AAM50644) contains a
XX Ca2+-dependent phospholipid binding domain and is associated with
XX the plasma membrane. BON1 belongs to the copine gene family, which
XX is conserved from protozoa to humans. The invention is directed to

CC isolated BON1, BON2, BON3, BAP1 and BAL nucleic acids (see
CC ABA91256-60), which encode proteins (see AAM50644-48) that are
CC necessary for normal growth, controlling cell expansion and cell
CC division, affecting the size and rate at which the plant grows when
CC exposed to lower temperatures. Transgenic plants are provided that
CC are smaller than the wild-type as a result of inhibition of BON1,
CC BON2, BON3, BAP1 and/or BAL, especially angiosperms and gymnosperms,
CC ornamental plants and turfgrass. Transgenic plants are also
CC provided that are larger than the wild-type as a result of
CC enhancement of BON1, BON2, BON3, BAP1 and/or BAL, especially crop
CC plants and biomass plants. Modulation of these genes provides
CC increased yield, or growth at a higher altitude or lower
CC temperature.

XX
SQ Sequence 2611 BP; 809 A; 429 C; 498 G; 866 T; 9 other;

Query Match 3.9%; Score 40.4; DB 24; Length 2611;
Best Local Similarity 97.6%; Pred. No. 0.074;
Matches 41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCGACGGTGTGTCGACGGCCCGGGCTGGTATG 42

Db 2603 CTTACTATAGGCGACGGTGTGTCGACGGCCCGGGCTGGTCTG 2562

RESULT 10

AAC62780

ID AAC62780 standard; DNA; 411 BP.

XX AAC62780;

XX 02-FEB-2001 (first entry)

XX Flower specific promoter coding sequence #2.

XX Promoter; eucalyptus; pine; gene transcription; ds.

XX Eucalyptus grandis.

XX WO200058474-A1.

XX 05-OCT-2000.

XX 24-FEB-2000; 2000WO-NZ000018.

XX 25-MAR-1999; 99US-0276599.

XX 30-JUL-1999; 99US-0146591.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX (FLET-) FLETCHER CHALLENGE FORESTS LTD.

XX Perera R, Rice SJ, Eagleton CK;

XX WPI; 2000-647236/62.

XX Novel promoter sequences useful for modulating transcription of plant
XX DNA sequences of interest and production of polypeptides

XX Claim 1; Page 53; 93pp; English.

XX The present invention relates to promoter sequences from eucalyptus
XX and pine. The present sequence is one such promoter. This sequence is
XX useful for modulating the transcription of DNA sequences of interest. The
XX sequences may also be used to tag or identify an organism or its
XX reproductive material.

XX Sequence 411 BP; 112 A; 118 C; 80 G; 101 T; 0 other;

Query Match 3.9%; Score 40.2; DB 21; Length 411;

Best Local Similarity 93.3%; Pred. No. 0.029; Length 411;

Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 ACTATAGGCGACGGTGTGTCGACGGCCCGGGCTGGTATGAAGGTG 48

```

Db      1  ACTATAGGCACGCGTGTGTCAGCGCCCGGCTGGTCTGAAACTG 45
|||||
RESULT 11
ABK17045
ID      ABK17045 standard; cDNA; 411 BP.
XX      AC      ABK17045;
XX      DT      26-MAR-2002 (first entry)
XX      DE      Eucalyptus grandis promoter polynucleotide #16.
XX      KW      Promoter; pine; leaf; root; flower; pollen; bud; meristem; xylogenesis;
KW      temporally regulated promoter; Pinus radiata; Eucalyptus grandis; ss;
KW      PCR primer.
XX      OS      Eucalyptus grandis.
XX      PN      WO200198485-A1.
XX      PD      27-DEC-2001.
XX      PF      20-JUN-2001; 2001WO-NF00115.
XX      PR      20-JUN-2000; 2000US-0598401.
XX      PR      28-NOV-2000; 2000US-0724624.
XX      PA      (GENE-) GENESIS RES & DEV CORP LTD.
XX      PA      (FLEET-) FLETCHER CHALLENGE FORESTS IND LTD.
XX      PI      Perera R, Rice S, Eagleton C, Lasham A;
XX      WI      WPI; 2002-114583/15.
XX      PT      Novel polynucleotide promoter sequences from Pine and Eucalyptus useful
PT      for modifying expression of endogenous and/or heterologous
PT      polynucleotides in transgenic plants
XX      PS      Claim 1; Page 68-69; 121pp; English.
XX      CC      The invention relates to isolated promoter sequences from Pinus radiata
CC      and Eucalyptus grandis, comprising a sequence chosen from leaf-, root-,
CC      flower-, pollen-, bud-, meristem-specific promoters or temporally
CC      regulated promoters such as xylogenesis-specific promoters. The promoter
CC      polypeptides and their related polynucleotides are useful in the
CC      production of genetic constructs, used for modifying gene expression in a
CC      target organism, in particular a plant. The method is useful for
CC      modifying expression of a polynucleotide that comprises an intron
CC      sequence, through removal of the intron sequence. The method is useful
CC      for modifying growth and development of plants, and cellular responses to
CC      external stimulus, such as environmental factors and disease pathogens.
CC      The sequences are useful in genome and physical mapping, in positional
CC      cloning of genes, in various assays to determine biological activity, to
CC      raise antibodies, to isolate corresponding interacting proteins and other
CC      compounds, and to quantitatively determine levels of interacting proteins
CC      or other compounds. Sequences ABK17016-ABK17125 represent Pinus radiata
CC      and Eucalyptus grandis polynucleotides and PCR primers used in the method
CC      of the invention.
XX      SQ      Sequence 411 BP; 112 A; 118 C; 80 G; 101 T; 0 other;

Query Match      3.9%; Score 40.2; DB 24; Length 411;
Best Local Similarity 93.3%; Pred. No. 0.029;
Matches 42; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY      4  ACTATAGGCACGCGTGTGTCAGCGCCCGGCTGGTATGAAGTG 48
|||||
Db      1  ACTATAGGCACGCGTGTGTCAGCGCCCGGCTGGTCTGAAACTG 45
|||||
RESULT 12

```

```

AAK35652
ID      AAK35652 standard; cDNA; 594 BP.
XX      AC      AAK35652;
XX      DT      09-JUL-1999 (first entry)
XX      DE      5' region of human heparanase cDNA.
XX      KW      Heparanase; hpa; modulator; heparin-binding growth factor;
KW      cellular response; cytokine; cell interaction; plasma lipoprotein;
KW      cellular susceptibility; infection; disintegration;
KW      neurodegenerative plaque; wound healing; angiogenesis; restenosis;
KW      atherosclerosis; inflammation; neurodegenerative disease; neutralise;
KW      plasma heparin; micrometastasis; autoimmune lesion; renal failure;
KW      ss.
XX      OS      Homo sapiens.
XX      PN      WO9911798-A1.
XX      PD      11-MAR-1999.
XX      PF      31-AUG-1998; 98WO-US17954.
XX      PR      02-JUL-1998; 98US-0109386.
XX      PR      02-SEP-1997; 97US-0922170.
XX      PA      (FRIE/) FRIEDMAN M M.
XX      PA      (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX      PA      (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX      PI      Feinstein E, Pecker I, Vlodaysky I;
XX      WI      WPI; 1999-302255/25.
XX      PT      New human polynucleotide useful for treating angiogenesis,
PT      restenosis, and inflammation
XX      PS      Example 8; Page 69; 63pp; English.
XX      CC      The specification describes a polypeptide having heparanase (hpa)
CC      activity. The recombinant protein is used as a modulator of
CC      heparin-binding growth factors, cellular responses to heparin-binding
CC      growth factors and cytokines, cell interaction with plasma lipoproteins,
CC      cellular susceptibility to viral, protozoal and bacterial infections
CC      or disintegration of neurodegenerative plaques. Heparanase may be
CC      useful for conditions such as wound healing, angiogenesis, restenosis,
CC      atherosclerosis, inflammation, neurodegenerative diseases, and viral
CC      infections. Mammalian heparanase can be used to neutralize plasma
CC      heparin, and anti-heparanase antibodies may be applied for
CC      immunodetection and diagnosis of micrometastases, autoimmune lesions,
CC      and renal failure in biopsy specimens, plasma samples, and body fluids.
CC      The present sequence represents the 5' sequence of human heparanase
CC      cDNA.
XX      SQ      Sequence 594 BP; 133 A; 134 C; 209 G; 118 T; 0 other;

Query Match      3.9%; Score 40; DB 20; Length 594;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2  TTACTATAGGCACGCGTGTGTCAGCGCCCGGCTGGTAT 41
|||||
Db      2  TTACTATAGGCACGCGTGTGTCAGCGCCCGGCTGGTAT 41
|||||
RESULT 13
AAK75054
ID      AAA75054 standard; cDNA; 594 BP.
XX      AC      AAA75054;
XX      CC

```

DT 15-JAN-2001 (first entry)
XX Upstream sequence of human cDNA encoding heparanase.
DE
XX
KW Human; heparanase; gene therapy; tumour; inflammation; autoimmunity;
KW heparin-binding growth factor; cytokine; neurodegenerative plaque;
KW wound healing; infection; burn; angiogenesis; restenosis;
KW atherosclerosis; inflammation; neurodegenerative disease;
KW Gerstmann-Straussler Syndrome; Creutzfeldt-Jakob disease; ds.
XX
OS Homo sapiens.
XX
XX WO200052178-A1.
XX
XX 08-SEP-2000.
XX
XX 14-FEB-2000; 2000WO-US03542.
XX
XX 01-MAR-1999; 99US-0258892.
XX
XX (INSI-) INSIGHT STRATEGY & MARKETING LTD.
XX (HADA-) HADASIT MEDICAL RES SERVICES & DEV.
XX (FRIE/) FRIEDMAN M M.
XX
XX Pecker I, Vlodaysky I, Feinstein E;
XX
XX WPI; 2000-579289/54.
XX
XX New polynucleotides encoding a polypeptide having heparanase activity,
XX useful in wound healing and in gene therapy, particularly in treating
XX tumour, inflammation, autoimmunity, neurodegenerative diseases -
XX
XX Example 7; Page 127; 152pp; English.
XX
XX The present sequence is an upstream sequence of cDNA encoding encodes a
XX human protein with heparanase catalytic activity. The heparanase (hpa)
XX polynucleotide is useful in gene therapy, particularly in treating
XX tumour, inflammation or autoimmunity. Particularly, the polynucleotide
XX is useful in modulating the bioavailability of heparin-binding growth
XX factors, cellular responses to heparin-binding growth factors (e.g. bFGF)
XX and cytokines (e.g. interleukin (Il)-8), cell interaction with plasma
XX lipoproteins, cellular susceptibility to certain viral and some bacterial
XX and protozoa infections, or disintegration of neurodegenerative plaques.
XX The polynucleotide is also useful in wound healing (e.g. thermal,
XX chemical or radiation burns), and in the treatment of angiogenesis,
XX restenosis, atherosclerosis, inflammation, neurodegenerative diseases
XX (Gerstmann-Straussler Syndrome or Creutzfeldt-Jakob disease), and some
XX viral, bacterial or protozoa infections.
XX
XX Sequence 594 BP; 133 A; 134 C; 209 G; 118 T; 0 other;
SQ
Query Match 3.9%; Score 40; DB 21; Length 594;
Best Local Similarity 100.0%; Pred. No. 0.041;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TTACTATAGGCGACGGTGTGTCGACGCCGCCGGCTGGTAT 41
|||||
DB 2 TTACTATAGGCGACGGTGTGTCGACGCCGCCGGCTGGTAT 41
RESULT 14
AAV08328
ID AAV08328 standard; cDNA; 1680 BP.
XX
XX AAV08328;
XX
XX 04-FEB-1999 (first entry)
XX
XX Hyaluronate synthase promoter.
DE
XX Hyaluronate synthase; promoter; human;
KW hyaluronic acid production inhibitor; ds.
XX

OS Homo sapiens.
XX
XX JP10295383-A.
XX
XX 10-NOV-1998.
XX
XX 24-APR-1997; 97JP-0107624.
XX
XX 24-APR-1997; 97JP-0107624.
XX
XX (ADSK-) ADVANCED SKIN RES KENKUSHO KK.
XX
XX WPI; 1999-038280/04.
XX
XX A promoter for hyaluronate synthase gene - used for screening of
XX modulators of the enzyme
XX
XX Claim 1; Page 5-6; 7pp; Japanese.
XX
XX This sequence represents the DNA of the invention, and can act as a
XX promoter for the hyaluronate synthase gene. It was isolated from human
XX cDNA sources. The DNA is useful for screening a drug enhancing or
XX inhibiting production of hyaluronic acid.
XX
XX Sequence 1680 BP; 452 A; 381 C; 467 G; 380 T; 0 other;
SQ
Query Match 3.9%; Score 40; DB 20; Length 1680;
Best Local Similarity 100.0%; Pred. No. 0.077;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTTACTATAGGCGACGGTGTGTCGACGCCGCCGGCTGGTA 40
|||||
DB 9 CTTACTATAGGCGACGGTGTGTCGACGCCGCCGGCTGGTA 48
RESULT 15
AAZ51547
ID AAZ51547 standard; DNA; 1224 BP.
XX
XX AAZ51547;
XX
XX 21-JUN-2000 (first entry)
XX
XX Seed-preferred promoter-2 from maize end gene.
XX
XX Seed-preferred promoter; maize; end gene; endosperm; end1; end2;
XX fatty acid; starch profile; carbohydrate profile; amino acid content;
XX agronomic; insect resistance; disease resistance; herbicide resistance;
XX grain characteristic; ds.
XX
XX Zea mays.
XX
XX WO200012733-A1.
XX
XX 09-MAR-2000.
XX
XX 25-AUG-1999; 99WO-US19604.
XX
XX 28-AUG-1998; 98US-0098230.
XX
XX (PION-) PIONEER HI-BRED INT INC.
XX
XX Martino-Catt SJ, Lappegard KK, Olsen O, Linnestad C, Abbitt SE;
XX
XX WPI; 2000-256648/22.
XX
XX Maize promoter driving transcription in a seed-preferred manner, for
XX stably transforming plant cells -
XX
XX Claim 1; Pages 35-36; 43pp; English.
XX
XX The present sequence is a seed-preferred promoter isolated from maize
XX endosperm specific genes, end1 or end2. The promoter is capable of
CC

CC driving transcription in a seed-preferred manner and can be used in an
CC expression cassette, to stably transform plant cells. The expression
CC cassette can be used to modify the fatty acid content of seeds, alter
CC the starch or carbohydrate profile, and/or alter the amino acid content
CC of the seed. It can also be used to deliver genes encoding important
CC traits for agronomics, including insect resistance, disease resistance,
CC herbicide resistance, and grain characteristics.

XX

SQ Sequence 1224 BP; 378 A; 239 C; 182 G; 425 T; 0 other;

Query Match 3.9%; Score 39.8; DB 21; Length 1224;
Best Local Similarity 95.3%; Pred. No. 0.074;
Matches 41; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 TACTATAGGCGACGCTGTGTCGACGCCGCCGGCTGGTATGAAG 45

|||||
1 TACTATAGGCGACGCTGTGTCGACGCCGCCGGCTGGTATGAAG 43

Search completed: December 28, 2002, 00:07:43
Job time : 254 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2002, 19:35:28 ; Search time 2673 Seconds
(without alignments)
11214.325 Million cell updates/sec

Title: US-09-945-376-3
Perfect score: 1030
Sequence: 1 cttactataggcagcggtg.....tgttatatacaaggtctgg 1030

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb.ba.*
2: gb.htg.*
3: gb.in.*
4: gb.om.*
5: gb.ov.*
6: gb.pat.*
7: gb.ph.*
8: gb.pl.*
9: gb.pr.*
10: gb.ro.*
11: gb.sts.*
12: gb.sy.*
13: gb.un.*
14: gb.vi.*
15: em.ba.*
16: em.fun.*
17: em.hum.*
18: em.in.*
19: em.mu.*
20: em.om.*
21: em.or.*
22: em.ov.*
23: em.pat.*
24: em.ph.*
25: em.pl.*
26: em.ro.*
27: em.sts.*
28: em.un.*
29: em.vi.*
30: em.htg_hum.*
31: em.htg_inv.*
32: em.htg_other.*
33: em.htg_mus.*
34: em.htg_pln.*
35: em.htg_rod.*
36: em.htg_mam.*
37: em.htg_vrt.*
38: em.sy.*
39: em.htgo_hum.*
40: em.htgo_mus.*
41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	49.8	4.8	125020	9	AF429315	AF429315 Homo sapi
C 2	44.8	4.3	125020	9	AF429315	AF429315 Homo sapi
C 3	44.4	4.3	216878	2	AL773534	AL773534 Mus muscu
C 4	42.8	4.2	685	8	AF457660	AF457660 Castanea
C 5	42.4	4.1	1423	9	AF039526	AF039526 Homo sapi
C 6	42.4	4.1	6078	6	AR124194	AR124194 Sequence
C 7	41.8	4.1	92633	2	PFMAL4P1_3	Continuation (4 of
C 8	41.8	4.1	110000	2	PFMAL4P1_2	Continuation (3 of
C 9	41	4.0	4260	5	AY124482	AY124482 Danio rer
C 10	41	4.0	11957	9	AF051769	AF051769 Homo sapi
C 11	40.8	4.0	1094	8	AF075270	AF075270 Hordium v
C 12	40.8	4.0	166703	9	AL139378	AL139378 Human DNA
C 13	40.6	3.9	156325	9	HS134N8	AL031655 Human DNA
C 14	40.4	3.9	598	6	AX320057	AX320057 Sequence
C 15	40.4	3.9	1463	3	AF153014	AF153014 Trichomon
C 16	40.4	3.9	2611	6	AX380786	AX380786 Sequence
C 17	40.2	3.9	2463	9	AB021922	AB021922 Homo sapi
C 18	40.2	3.9	22031	9	AC025822	AC025822 Homo sapi
C 19	40	3.9	498	10	RNU53907	U53907 Rattus norv
C 20	40	3.9	951	3	AF233737	AF233737 Agrotis i
C 21	40	3.9	985	6	AX253432	AX253432 Sequence
C 22	40	3.9	1499	6	AX253428	AX253428 Sequence
C 23	40	3.9	1958	10	AF114032	AF114032 Mus muscu
C 24	39.8	3.9	682	10	MMU289605	AJ289605 Mus muscu
C 25	39.8	3.9	2385	6	AX041981	AX041981 Sequence
C 26	39.8	3.9	349980	6	AX344560	AX344560 Sequence
C 27	39.6	3.8	7218	6	I66494	I66494 Sequence 14
C 28	39.4	3.8	640	10	MMU63899	U63899 Mus musculu
C 29	39.4	3.8	1135	9	AF074397	AF074397 Homo sapi
C 30	39.4	3.8	2492	6	AX098431	AX098431 Sequence
C 31	39.4	3.8	2867	6	E26792	E26792 Novel prote
C 32	39.4	3.8	4185	1	AF237414	AF237414 Ehrlichia
C 33	39	3.8	516	6	AX320056	AX320056 Sequence
C 34	39	3.8	615	10	AF077743	AF077743 Mus muscu
C 35	39	3.8	2184	6	AX253431	AX253431 Sequence
C 36	39	3.8	2785	10	AF162890S1	AF162890 Mus muscu
C 37	38.8	3.8	104526	9	AC087887	AC087887 Homo sapi
C 38	38.8	3.8	149440	2	AC024319	AC024319 Homo sapi
C 39	38.8	3.8	196049	2	AC105841	AC105841 Rattus no
C 40	38.8	3.8	299350	1	SME591786	AL591786 Sinorhizo
C 41	38.8	3.8	329100	1	SME591787	AL591787 Sinorhizo
C 42	38.6	3.7	1157	8	AF457661	AF457661 Castanea
C 43	38.6	3.7	5031	6	AX085162	AX085162 Sequence
C 44	38.6	3.7	5031	6	AX085359	AX085359 Sequence
C 45	38.6	3.7	180968	2	AC078996	AC078996 Mus muscu

ALIGNMENTS

RESULT 1
AF429315/c
LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION AF429315
ACCESSION AF429315
VERSION AF429315.1 GI:17645244
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.

TITLE A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2

JOURNAL Nat. Genet. 29 (4), 377-378 (2001)

MEDLINE 21583737

PUBMED 11694876

REFERENCE 2 (bases 1 to 125020)

AUTHORS Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.

TITLE Direct Submission

JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA

FEATURES

Location/Qualifiers

1. 125020

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/map="16q24.3; between D16S520 and WI-12410"

/note="Isolated from a patient with Huntington's Disease-Like 2 (HDL2)"

complement(35581..35746)

/rpt_type=tandem

/rpt_unit=ctg

complement(<36507..>36887)

/gene="JPH3"

/note="JP3"

complement(<36507..>36887)

/gene="JPH3"

/product="junctophilin 3"

complement(<36507..36887)

/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"

/codon_start=1

/product="junctophilin 3"

/protein_id="AAL40941.1"

/db_xref="GI:17646245"

/translation="MSSGGRNFDDGSGYCGWEDGKAHGHCVGTGPKGQGYTGWS HGFEVLGYTPSGNYGTQWAKRHRHIGLESKGKVVYKGEWTHGKRGYGVRECA NGAKYEGTWSNGLQDGYGTETYS DG"

BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others

ORIGIN

Query Match 4.8%; Score 49.8; DB 9; Length 125020;

Best Local Similarity 9.3%; Pred. No. 0.007;

Matches 41; Conservative 213; Mismatches 185; Indels 1; Gaps 1;

QY 525 AGCCCTGAACCGCTGCTACTGCTTC-CGCGCGGACCGGACCTCCTCCTGCTGCTGCCA 703

Db 17666 WSSYSTSMGYSTCKYKCSWSSMYKCKTSKYRRKRSYIYWGGRKAKKYYCAGRR 17607

QY 585 GAACTTGCTCAACTCTGCATTTTCGGTTTCAGACAGTACTGCAATTTTGCCATTGTCG 644

Db 17606 RRMVYWKCAKWMVSYCCWYSYCMYTYYSKSTCYKRSCTCYKRGYYWGSKTCTYSAGGKRS 17547

QY 645 CTGCACGACGACTGCTTACTGCTTC-CGCGCGGACCGGACCTCCTCCTGCTGCTGCCA 703

Db 17546 MYCYMRSSKSSWSMWSMARSWCMGWAGYRRSRAGWAGWRSXGKRSTGKMRKS 17487

QY 704 ACTGCCAACACGACCTGCATCGCCTGGTGGCGTCATCTAATTTCTCTATCAAGAATC 763

Db 17486 KKTGSYGTGRSMKKKGYSKYSRGMKCKKCTCYCMWVYKYKRTSMCWVYMKSWGYK 17427

QY 764 ACCTCCCATTTATCCGCCCTGCGCGGACCTCGCTCCCTCAATTCACGGGCACACGG 823

Db 17426 YKRCMKKKGCTGYKMSKSKSGYKSMRMRGSSYSTSCWKSCHWYKMKCMYKSK 17367

QY 824 TGCAGGCACACCGCCCATCAAAATACGGCTCTTTTCGGCAGCATGCTCCACCGACGG 883

Db 17366 RSRMRGSMSSKGRWYAGRGYSSSWSTRKRSKCYKSKYKGRKMKMGWGMKRGSK 17307

QY 884 CGCCTAATGAATTCAGCTGATGAGTGGTGGATCGCGAGGAAACCTTGAATGGC 943

Db 17306 YWSSMKMKRRSSWRCYSTKYKSKRRSKGWRSTKSKAKSSMRMAGSKCTCTGSSYSN 17247

QY 944 CGGCATCAGTAGCACCCGCCA 963

Db 17246 NRRNRMGKTCGNYMRSR 17227

RESULT 2

AF429315

LOCUS Homo sapiens junctophilin 3 (JPH3) gene, partial cds.

DEFINITION AF429315

ACCESSION AF429315

VERSION AF429315.1 GI:17646244

KEYWORDS

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 125020)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S., Ingersoll-Ashworth, R.G., Fleisher, A., Stevanin, G., Brice, A., Potter, N.T., Ross, C.A. and Margolis, R.L.

TITLE A repeat expansion in the gene encoding junctophilin-3 is associated with Huntington disease-like 2

JOURNAL Nat. Genet. 29 (4), 377-378 (2001)

MEDLINE 21583737

PUBMED 11694876

REFERENCE 2 (bases 1 to 125020)

Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.

TITLE Direct Submission

JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA

FEATURES

Location/Qualifiers

1. 125020

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/map="16q24.3; between D16S520 and WI-12410"

/note="Isolated from a patient with Huntington's Disease-Like 2 (HDL2)"

complement(35581..35746)

/rpt_type=tandem

/rpt_unit=ctg

complement(<36507..>36887)

/gene="JPH3"

/note="JP3"

complement(<36507..>36887)

/gene="JPH3"

/product="junctophilin 3"

complement(<36507..36887)

/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"

/codon_start=1

/product="junctophilin 3"

/protein_id="AAL40941.1"

/db_xref="GI:17646245"

/translation="MSSGGRNFDDGSGYCGWEDGKAHGHCVGTGPKGQGYTGWS HGFEVLGYTPSGNYGTQWAKRHRHIGLESKGKVVYKGEWTHGKRGYGVRECA NGAKYEGTWSNGLQDGYGTETYS DG"

BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others

ORIGIN

Query Match 4.3%; Score 44.8; DB 9; Length 125020;

Best Local Similarity 10.1%; Pred. No. 0.22;

Matches 53; Conservative 260; Mismatches 207; Indels 4; Gaps 1;

QY 470 CAGACGAGACAAATTCGGCATCAACACCTTTTCGGCCACAACTATGTTGGTAGGCC 529

Db 50579 CTGRWTGKBCKSSKSHRRRTMTDYMRKYTYCCCTATYMBBYHYCDBCSWGY 50638

QY 530 TGAACCGCTCGCTATTCAGTTGGGTGGTGGCCACCGCATGATGCCATCCCAACT 589

Db 50639 GGWRGMMKMBWBBSBMGKSMGDTDKDYWCYVSSMSVDVYVWRWBDTSWGSBSHYR 50698

QY 590 TGCTCCAACTGTCATTTGGTTCAGACAG-----TACTGCATTTGGCCATTTGTCGCC 645
Db 50699 GVHTGVRWVGBCDTHVTYDGHHSWGMKSCGYSDBGCAKBSKMTGTSFKYTC 50758
QY 646 TGCACGAGACTGCTTACTGTTCCGGCGGGAACCGGACCTCCCTCCTGCTCTGCCAAC 705
Db 50759 YGKYWSSKADBSYYBSHBWGSVMCCRKBSYASSYDVRYKBTGMYRVRSWY 50818
QY 706 TGCACACGAGCTGCATGGCTGGTGGCGTCACTAATTTCTCTCATCAAGATCATC 765
Db 50819 YVKYBSMSVSRGSKGHKVRBSMKCSRRKDMSTSSMWRSMKDRCKSHSHYSKSMGK 50878
QY 766 CTCCCATTTACTCGCCCTGTGGCGGACCTCGTGCCTCGAATTCACGCGGCACACGGTG 825
Db 50879 WTSDDVRCVACSBMYKMKRKHKKHDSBWSKSHMKKGAWRVSRMYSRCSVHDSS 50938
QY 826 CAGGCACACACCGCCGCAATAATAGCGCTCTTTGGCGGACGGATCTCCACCGCAGGGC 885
Db 50939 MVDKSRMRKSGAMRMKCYCTTSSMRSTKSRSMKSSWRGSKYCYGYWSTCKMRSMG 50998
QY 886 GCAATTGAATTCAGCTGATCGAGTGGTGGATCGCGCAGGAAACCTTGAATGGCGG 945
Db 50999 CSYSTGSSMKMGKYCCWAGSSRYCKSGSRMSWAGSSMSMGKMTSGMSGRCCA 51058
QY 946 GCATAGTAGCAGCGCAGCTCACCTCTCGAAGAAGAGCTGTGCG 989
Db 51059 RSGMSAYCSCRCKSKRCSGCMRCWYCMRYGSMRYKC 51102

AL773534 216878 bp DNA linear HTG 17-AUG-2002
Mus musculus chromosome 2 clone RP23-183012, *** SEQUENCING IN
PROGRESS ***, 2 unordered pieces.
AL773534
HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Johnson,C.
Direct Submission
Submitted (16-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Aug 21, 2002 this sequence version replaced gi:22002783.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bM183012
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: dye-terminator; 100% of reads
Consensus quality: 216691 bases at least Q40
Consensus quality: 216754 bases at least Q30
Consensus quality: 216766 bases at least Q20
Insert size: 216778; sum-of-contigs
Insert size: 211070; 5.8% error; agarose-fp
Quality coverage: 7.01x in Q20 bases; sum-of-contigs Quality
coverage: 7.25x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will

* be preserved.
1 15291: contig of 15291 bp in length
* 15292 15391: gap of 100 bp
* 15392 216878: contig of 201487 bp in length.
FEATURES
source
Location/Qualifiers
1..216878
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="2"
/clone="RP23-183012"
/clone_lid="RPC1-23"
1..15291
/note="assembly_fragment:02090
fragment_chain:1"
15392..216878
/note="assembly_fragment:03034
fragment_chain:1"
BASE COUNT 70260 a 41471 c 41503 g 63544 t 100 others
ORIGIN
Query Match 4.38; Score 44.4; DB 2; Length 216878;
Best Local Similarity 47.58; Pred. No. 0.31;
Matches 132; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 177 AAGATGTGGATTACACAGTTCTATCTAGCGACAGTCTAAACGACGCTTTGTATAGTC 236
Db 137525 AATTGCTCTAAAATAGCAAAATTTACAAGATTGACTATATACTTCCCTCTGTATTAGTT 137584
QY 237 ATTTTATCTCCCAACCGCATTTAGTAAGCAGGCCCCCAACGCAATTCAAATTCGATCTA 296
Db 137585 ATTTGCATTTGATGAACGTCTTTATTTTAAATGATTAGAAATAGTACATTTAAAGGA 137644
QY 297 ATTTGTCTACTTCATTTTGGTTCAGATTTGAAGATGTTCAGTTTCAATCCACATGCAACAG 356
Db 137645 CTTATCATATTGCTCTCTCTTTATTTACTAGATATCATTTAATTTCAATCCAGCCA 137704
QY 357 AATCCGACAGAGATTCATCAATTCACACATAAAGACAGCAAGAGTTAAACCATAAA 416
Db 137705 ATTTGACTATAGATATTTTCAGATTCAAATTAATCAAGACATCAGTAAAGTATAAAA 137764
QY 417 CATCAATTAAGGATTCGATTTTAGATCTTTTAGAGA 454
Db 137765 CATCATGAATTTGGTATTAGATGCTGCTTTTCATAGA 137802
RESULT 4
AF457660 685 bp DNA linear PLN 13-JAN-2002
LOCUS Castanea dentata clone ACS2 vascular protein gene, promoter region.
DEFINITION AF457660
ACCESSION AF457660
VERSION AF457660.1 GI:18140057
KEYWORDS
SOURCE Castanea dentata.
ORGANISM Castanea dentata
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fagales; Fagaceae; Castanea.
REFERENCE
1 (bases 1 to 685)
Connors,B.J., Maynard,C.A. and Powell,W.A.
Cloning and characterization of promoters from American chestnut
capable of directing reporter gene expression in transgenic
Arabidopsis plants
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 685)
Connors,B.J., Maynard,C.A. and Powell,W.A.
AUTHORS Direct Submission
TITLE Submitted (12-DEC-2001) EPB, SUNY-ESF, 1 Forestry Drive, Syracuse,
JOURNAL NY 13210-2788, USA
FEATURES
Location/Qualifiers
1..685
/organism="Castanea dentata"
/db_xref="taxon:134033"
/clone="ACS2"

```
misc_feature 1..695
/Note="vascular protein promoter region"
BASE COUNT 253 a 126 c 117 g 189 t
ORIGIN
Query Match 4.2%; Score 42.8; DB 8; Length 685;
Best Local Similarity 63.7%; Pred. No. 0.31; Mismatches 37; Indels 0; Gaps 0;
Matches 65; Conservative 0;
Qy 1 CTTACTATAGGCACGGCTGTCAGCGCGGCTGGTATGAAGTGGGAACCTCACTG 60
Db 8 CTTACTATAGGCACGGCTGTCAGCGCGGCTGGTATGAAGTGGGAACCTCACTG 67
Qy 61 GATGCATATACCTGCTGAGAGATAACATCAACAATTCACAAGCT 102
Db 68 AAGATAAAAAATATTTAAACTGTAAGATAAAAAATAAAGCT 109
RESULT 5
AF039526 1423 bp DNA linear PRI 01-MAY-2001
LOCUS Homo sapiens MHC class I related protein 1 (MRL) gene, partial cds.
ACCESSION AF039526
VERSION AF039526.1 GI:4104807
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1423)
AUTHORS Riegert,P., Wanner,V. and Bahram,S.
TITLE Genomics, isoforms, expression, and phylogeny of the MHC class
I-related MRL gene
JOURNAL J. Immunol. 161 (8), 4066-4077 (1998)
MEDLINE 98451457
PUBMED 9780177
REFERENCE 2 (bases 1 to 1423)
AUTHORS Riegert,P., Wanner,V., Hauptmann,G. and Bahram,S.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1997) EG, Basel Institute for Immunology,
Grenzacherstrasse 487, Basel 4005, Switzerland
FEATURES
Location/Qualifiers
1..1423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="lq25"
1..>1423
/gene="MRL"
/misc_feature 1..1389
/gene="MRL"
/Note="contains putative promoter and 5'UTR"
<1390..>1423
/gene="MRL"
/product="MHC class I related protein 1"
/evidence=not_experimental
1390..>1423
/gene="MRL"
/codon_start=1
/evidence=not_experimental
/product="MHC class I related protein 1"
/protein_id="AAD02172.1"
/db_xref="GI:4104808"
/craniation="MGELMAFLPL"
342 a 334 c 310 g 437 t
BASE COUNT
ORIGIN
Query Match 4.1%; Score 42.4; DB 9; Length 1423;
Best Local Similarity 97.7%; Pred. No. 0.47; Mismatches 1; Indels 0; Gaps 0;
Matches 43; Conservative 0;
Qy 1 CTTACTATAGGCACGGCTGTCAGCGCGGCTGGTATGAAGTGGGAACCTCACTG 44
|||||
Db 1 CTTACTATAGGCACGGCTGTCAGCGCGGCTGGTATGAAGTGGGAACCTCACTG 60
|||||
RESULT 6
AF039526 1423 bp DNA linear PRI 01-MAY-2001
LOCUS Homo sapiens MHC class I related protein 1 (MRL) gene, partial cds.
ACCESSION AF039526
VERSION AF039526.1 GI:4104807
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1423)
AUTHORS Riegert,P., Wanner,V. and Bahram,S.
TITLE Genomics, isoforms, expression, and phylogeny of the MHC class
I-related MRL gene
JOURNAL J. Immunol. 161 (8), 4066-4077 (1998)
MEDLINE 98451457
PUBMED 9780177
REFERENCE 2 (bases 1 to 1423)
AUTHORS Riegert,P., Wanner,V., Hauptmann,G. and Bahram,S.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1997) EG, Basel Institute for Immunology,
Grenzacherstrasse 487, Basel 4005, Switzerland
FEATURES
Location/Qualifiers
1..1423
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/map="lq25"
1..>1423
/gene="MRL"
/misc_feature 1..1389
/gene="MRL"
/Note="contains putative promoter and 5'UTR"
<1390..>1423
/gene="MRL"
/product="MHC class I related protein 1"
/evidence=not_experimental
1390..>1423
/gene="MRL"
/codon_start=1
/evidence=not_experimental
/product="MHC class I related protein 1"
/protein_id="AAD02172.1"
/db_xref="GI:4104808"
/craniation="MGELMAFLPL"
342 a 334 c 310 g 437 t
BASE COUNT
ORIGIN
Query Match 4.1%; Score 42.4; DB 9; Length 1423;
Best Local Similarity 97.7%; Pred. No. 0.47; Mismatches 1; Indels 0; Gaps 0;
Matches 43; Conservative 0;
Qy 1 CTTACTATAGGCACGGCTGTCAGCGCGGCTGGTATGAAGTGGGAACCTCACTG 44
|||||
Db 1 CTTACTATAGGCACGGCTGTCAGCGCGGCTGGTATGAAGTGGGAACCTCACTG 60
|||||
RESULT 7
PFMAL4PL_3/c
WPCOMMENT
Sequence split into 4 fragments LOCUS PFMAL4PL Accession AL034557
Fragment Name Begin End
PFMAL4PL_0 1 110000
PFMAL4PL_1 100001 210000
PFMAL4PL_2 200001 310000
PFMAL4PL_3 300001 392633
Continuation (4 of 4) of PFMAL4PL from base 300001 (AL034557 Plasmidium falciparum 3D
Query Match 4.1%; Score 41.8; DB 2; Length 92633;
Best Local Similarity 65.6%; Pred. No. 1.6; Mismatches 0; Gaps 0;
Matches 61; Conservative 0;
Qy 1 CTTACTATAGGCACGGCTGTCAGCGCGGCTGGTATGAAGTGGGAACCTCACTG 60
|||||
Db 7315 CTTACTATAGGCACGGCTGTCAGCGCGGCTGGTATGAAGTGGGAACCTCACTG 7256
Qy 61 GATGCATATACCTGCTGAGAGATAACATCAACAAT 93
|||||
Db 7255 TTTCATATTTAAAAAAGGCTCATATAAAT 7223
RESULT 8
PFMAL4PL_2/c
WPCOMMENT
Sequence split into 4 fragments LOCUS PFMAL4PL Accession AL034557
Fragment Name Begin End
PFMAL4PL_0 1 110000
PFMAL4PL_1 100001 210000
PFMAL4PL_2 200001 310000
PFMAL4PL_3 300001 392633
Continuation (3 of 4) of PFMAL4PL from base 200001 (AL034557 Plasmidium falciparum 3D
Query Match 4.1%; Score 41.8; DB 2; Length 110000;
Best Local Similarity 65.6%; Pred. No. 1.6; Mismatches 0; Gaps 0;
Matches 61; Conservative 0;
Qy 1 CTTACTATAGGCACGGCTGTCAGCGCGGCTGGTATGAAGTGGGAACCTCACTG 60
|||||
Db 1 CTTACTATAGGCACGGCTGTCAGCGCGGCTGGTATGAAGTGGGAACCTCACTG 60
|||||
```

|||||
Db 107315 CTTACTATAGGCACGCGTGGTCGACGCCGCGCTGGTGAATATTTTTTTTTTTTTTTT 107256
QY 61 GATGCATATCTGCTGAGAGATAACATCAAT 93
|||||
Db 107255 TTCCCATATTTAAAAAGGCTCATATAAT 107223
RESULT 9
AY124482/c 4260 bp DNA linear VFT 31-JUL-2002
LOCUS Danio rerio myogenin gene, exons 1, 2 and 3 and complete cds.
DEFINITION AY124482
ACCESSION AY124482
VERSION AY124482.1 GI:22035404
KEYWORDS
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 4260)
Du,S., Gao,J. and Anyangwe,V.
Muscle-specific expression of myogenin in zebrafish embryos is
controlled by multiple regulatory elements in the promoter
Unpublished
JOURNAL
REFERENCE 2 (bases 1 to 4260)
AUTHORS Du,S., Gao,J. and Anyangwe,V.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Center of Marine Biotechnology, University
of Maryland Biotechnology Institute, 701 E. Pratt St, Baltimore, MD
21202, USA
FEATURES
source Location/Qualifiers
1..4260
/organism="Danio rerio"
/db_xref="taxon:7955"
promoter 1..651
mRNA join(652)..1309,2288..2393,3002..3599)
exon 652..1309
/product="myogenin"
/number=1
CDS join(764)..1309,2288..2393,3002..3120)
/codon_start=1
/product="myogenin"
/protein_id="AA082616.1"
/db_xref="GI:22035405"
/translation="MELFETNPFFNDQRFYEGADNFQSRINGFEQAGYQDRNSMM
GLCGDRLMTTVGLDEKPSFSSLSGMSPHQEQHCPCQCLPWACKVKRKSVTMD
RRKAATLRKRLKRVNEAFELKRSITLMPNQRILPKVILRLSAIQYIERLQALVSSL
NOOHEQGNLHYRTAAAPHTGVSSSDQSGTCCSPSEWSSASDHCVPAYSSAHED
LNDSSSQSNLRLSTSIVDSITGTPTVAYSVDISK"
2288..2393
/number=2
exon 3002..3599
/number=3
BASE COUNT 1361 a 797 c 731 g 1371 t
ORIGIN
Query Match 4.0%; Score 41; DB 5; Length 4260;
Best Local Similarity 89.8%; Pred. No. 1.5;
Matches 44; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 TTACTATAGGCACGCGTGGTCGACGCCGCGCTGTATGAGGTGG 50
|||||
Db 4242 TTACTATAGGCACGCGTGGTCGACGCCGCGCTGTGTAAGTACGAGG 4194
RESULT 10
AF051769/c 11957 bp DNA linear PRI 31-OCT-1999
LOCUS Homo sapiens hyaluronidase-like pseudogene 1 (HYALP1), partial
DEFINITION AF051769
sequence.
ACCESSION AF051769
VERSION AF051769.1 GI:6164589

KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 11957)
Csoka,A.B., Scherer,S.W. and Stern,R.
Expression analysis of six paralogous human hyaluronidase genes
clustered on chromosomes 3p21 and 7q31
Genomics 60 (3), 356-361 (1999)
JOURNAL
MEDLINE 99425275
PUBMED 10493834
REFERENCE 2 (bases 1 to 11957)
AUTHORS Csoka,T.B.
TITLE Direct Submission
JOURNAL Submitted (28-FEB-1998) Pathology, University of California,
Parnassus St. at 3rd Ave, San Francisco, CA 94143, USA
FEATURES
source Location/Qualifiers
1..11957
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q31.3"
<1..>11957
/gene="HYALP1"
/note="hyaluronidase-like pseudogene 1; exon/intron
structure not yet determined"
/pseudo
BASE COUNT 3424 a 2244 c 2085 g 4202 t 2 others
ORIGIN
Query Match 4.0%; Score 41; DB 9; Length 11957;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CTTACTATAGGCACGCGTGGTCGACGCCGCGGTGGTAT 41
|||||
Db 11955 CTTACTATAGGCACGCGTGGTCGACGCCGCGGTGGTAT 11915
RESULT 11
AF075270 1094 bp DNA linear PLN 24-SEP-1998
LOCUS Hordeum vulgare high affinity sulfate transporter (HVST1) gene,
DEFINITION promoter region, 5'UTR, and partial cds.
ACCESSION AF075270
VERSION AF075270.1 GI:3643823
KEYWORDS
SOURCE Hordeum vulgare subsp. vulgare.
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Hordeum.
1 (bases 1 to 1094)
Vidmar,J.J., Schjoerring,J.K., Touraine,B. and Glass,A.M.D.
Regulation of the HVST1 gene encoding a high affinity sulfate
transporter from Hordeum vulgare
Unpublished
JOURNAL 2 (bases 1 to 1094)
REFERENCE Vidmar,J.J., Schjoerring,J.K. and Glass,A.M.D.
AUTHORS Direct Submission
TITLE Submitted (30-JUN-1998) Botany, University of British Columbia,
JOURNAL 6270 University Blvd., Vancouver, B.C. V6T1Z4, Canada
FEATURES
source Location/Qualifiers
1..1094
/organism="Hordeum vulgare subsp. vulgare"
/cultivar="Klondike"
/db_xref="taxon:112509"
1..>1094
/gene="HVST1"
1..1024
/gene="HVST1"
1025..>1094
gene
promoter
mRNA

5'UTR /gene="HVS11"
1025..1089 /product="high affinity sulfate transporter"
CDS /gene="HVS11"
1090..>1094 /codon_start=1
/product="high affinity sulfate transporter"
/protein_id="AAC42664.1"
/db_xref="GI:3643824"
/translation="WP"

BASE COUNT 286 a 261 c 243 g 304 t
ORIGIN

Query Match 4.0%; Score 40.8; DB 8; Length 1094;
Best Local Similarity 95.5%; Pred. No. 1.3;
Matches 42; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCACGGCTGTCGACGCCGGCTGGTATGAA 44
Db 63 CTTACTATAGGCACGGCTGTCGACGCCGGCTGGTATGCA 106

RESULT 12
AL139378/c

LOCUS ALL139378 166703 bp DNA linear PRI 02-DEC-2000
DEFINITION Human DNA sequence from clone RP11-271B5 on chromosome 13 Contains
a gene for a protein similar to ribosomal protein S7, the FGF9
(fibroblast growth factor 9 (glia-activating factor)) gene, ESTs,
STSS, GSSs and CpG islands, complete sequence.

ACCESSION AL139378
VERSION AL139378.15 GI:8247510
KEYWORDS HTG; CpG island; FGF9; growth factor; ribosomal protein.
SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 166703)

AUTHORS Blakey,S.

TITLE Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
JOURNAL CB10 ISA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonesrequest@sanger.ac.uk

COMMENT On Jun 4, 2000 this sequence version replaced gi:8246895.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.

The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13

RP11-271B5 is from the library RPC1-11.1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://pacpac.med.buffalo.edu/
VECTOR: pBACE3.6

THIS sequence is the entire insert of clone RP11-271B5 The true
left end of clone RP11-282A24 is at 82919 in this sequence. The
true right end of clone RP11-110K8 is at 72244 in this sequence.
Location/Qualifiers

source 1..166703
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-271B5"
/clone_lib="RPC1-11.1"
27..485
/note="match: GSS: Em:B83261"
463..580
/note="MER5A repeat: matches 17..138 of consensus"
1191..1400
/note="MER3 repeat: matches 1..209 of consensus"
1669..1922
/note="AluJb repeat: matches 1..294 of consensus"
2320..2372
/note="12 repeat: matches 2700..2750 of consensus"
2621..2932
/note="MLTIC repeat: matches 147..466 of consensus"
2933..3198
/note="L1PA16 repeat: matches 5853..6157 of consensus"
3250..3410
/note="MLTIC repeat: matches 1..151 of consensus"
4184..4319
/note="12 repeat: matches 2609..2748 of consensus"
4703..4947
/note="12 repeat: matches 1807..2082 of consensus"
5016..5305
/note="AluSx repeat: matches 1..288 of consensus"
5493..5804
/note="AluY repeat: matches 1..311 of consensus"
6576..6869
/note="AluSc repeat: matches 1..289 of consensus"
7077..7157
/note="NSTC repeat: matches 291..359 of consensus"
7324..7630
/note="AluY repeat: matches 1..304 of consensus"
7766..7889
/note="FLAM_C repeat: matches 11..132 of consensus"
8776..9085
/note="AluSx repeat: matches 5..312 of consensus"
9516..9818
/note="AluY repeat: matches 1..303 of consensus"
10023..10311
/note="AluY repeat: matches 1..293 of consensus"
complement(10938..11116)
/note="match: GSS: Em:AQ384546"
complement(10938..11050)
/note="match: GSS: Em:AQ219188"
complement(10953..11555)
/gene="BA271B5.1"
complement(10953..11555)
/gene="BA271B5.1"
/note="match: cDNAs: Em:M77233
match: ESTs: Em:AAL05661 Em:AI462024
match: proteins: Tr:Q92NS1 Sw:Q10101 Sw:O43105 Tr:Q9VA91
Tr:Q23312 Sw:P02362 Sw:P23821 Sw:P48155 Sw:P33514
Tr:Q29190 Tr:O13167 Sw:P50894"
/codon_start=1
/evidence="not experimental"
/product="BA271B5.1 (similar to ribosomal protein S7)"
/protein_id="CAC17691.1"
/db_xref="GI:1154936"
/translation="MFSSSAKIVKPNDEKDFEFGISQALLEMNSDKLAQLRELN
ITAAKEIELGGGRKAIIFVPIPOLKSFQIOVLVRELEKFGKHVVFIIVORITLP
KPTRKSRKTKOKRPRSHLTAVHDAILEDLVPSVEIGKRIRVKLDGSLRIKVLHDK
AQQNVHKEVFSGYVKLTGKDVFNEPEPEQLQTKMT"
complement(10958..11260)
/gene="BA271B5.1"
/note="match: GSS: Em:AQ887379"
10995..11131
/note="match: GSS: Em:AQ476700"
10997..11430
/note="match: GSS: Em:AQ593042"

misc_feature

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 27, 2002, 19:37:43 : Search time 1965 seconds
(without alignments)
8489.237 Million cell updates/sec

Title: US-09-945-376-3
Perfect score: 1030
Sequence: 1 ctctatcaggcgcgctg.....tgtatatatacaaggctctgg 1030

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: em_estba.*
2: em_esthum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_vrt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	45	4.4	388	17 BH747511	BH747511 SALK_0180
c 2	43.6	4.2	469	17 BH251364	BH251364 SALK_0114
c 3	43.4	4.2	111	17 BH617520	BH617520 SALK_0372
c 4	42.8	4.2	124	17 BH613783	BH613783 SALK_0349
c 5	42.8	4.2	384	17 BH250931	BH250931 SALK_0107
c 6	42.8	4.2	445	17 BH634101	BH634101 SALK_0447

c 7	42.6	4.1	440	17 BH748887	BH748887 SALK_0469
c 8	42.2	4.1	432	17 BH610787	BH610787 SALK_0180
c 9	42.2	4.1	500	17 BH214178	BH214178 SALK_0103
c 10	42.2	4.1	1101	17 CNS001G0	AL068402 Drosophila
c 11	42	4.1	129	17 BH168812	BH168812 SALK_0002
c 12	42	4.1	220	17 BH251572	BH251572 SALK_0118
c 13	42	4.1	452	17 BH212774	BH212774 SALK_0080
c 14	41.8	4.1	416	17 BH749138	BH749138 SALK_0473
c 15	41.6	4.0	457	17 BH172739	BH172739 SALK_0061
c 16	41.6	4.0	471	17 BH251074	BH251074 SALK_0109
c 17	41.6	4.0	502	17 BH253202	BH253202 SALK_0146
c 18	41.4	4.0	120	17 BH750440	BH750440 SALK_0408
c 19	41.4	4.0	430	17 BH612486	BH612486 SALK_0328
c 20	41.4	4.0	445	17 BH612500	BH612500 SALK_0328
c 21	41.4	4.0	456	17 BH617406	BH617406 SALK_0364
c 22	41.4	4.0	456	17 BH633739	BH633739 SALK_0431
c 23	41.2	4.0	404	17 BH617295	BH617295 SALK_0362
c 24	41.2	4.0	465	17 BH170645	BH170645 SALK_0031
c 25	41.2	4.0	469	17 BH749118	BH749118 SALK_0473
c 26	41	4.0	235	17 BH251060	BH251060 SALK_0109
c 27	41	4.0	248	17 BH254262	BH254262 SALK_0162
c 28	41	4.0	455	17 BH171573	BH171573 SALK_0045
c 29	40.8	4.0	433	17 BH610810	BH610810 SALK_0180
c 30	40.8	4.0	451	17 BH254421	BH254421 SALK_0165
c 31	40.8	4.0	477	17 BH170269	BH170269 SALK_0026
c 32	40.6	3.9	105	17 BH752334	BH752334 SALK_0140
c 33	40.6	3.9	121	17 BH616998	BH616998 SALK_0358
c 34	40.6	3.9	128	17 BH752369	BH752369 SALK_0184
c 35	40.6	3.9	218	17 BH170998	BH170998 SALK_0036
c 36	40.6	3.9	266	17 BH611050	BH611050 SALK_0297
c 37	40.6	3.9	454	17 BH611042	BH611042 SALK_0297
c 38	40.6	3.9	456	17 BH252514	BH252514 SALK_0135
c 39	40.6	3.9	492	17 BH611714	BH611714 SALK_0315
c 40	40.4	3.9	438	17 BH251902	BH251902 SALK_0123
c 41	40.4	3.9	460	17 BH188842	BH188842 036_P_10-
c 42	40.4	3.9	460	17 CNS07SRN	AL625781 T3 end of
c 43	40.4	3.9	470	17 BH172612	BH172612 SALK_0059
c 44	40.2	3.9	114	17 BH250920	BH250920 SALK_0107
c 45	40.2	3.9	116	17 BH251621	BH251621 SALK_0118

ALIGNMENTS

RESULT 1
BH747511/c 388 bp DNA linear GSS 27-FEB-2002
LOCUS
DEFINITION
SALK_018023.45.85.x Arabidopsis thaliana TDNA insertion lines
Arabidopsis thaliana genomic clone SALK_018023.45.85.x, DNA
sequence.
BH747511 GI:18960626
BH747511
GSS.
SOURCE
Arabidopsis thaliana
thale cress.
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 388)
REFERENCE
Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
, Zimmerman,J. and Ecker,J.R.
TITLE
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome
JOURNAL
Unpublished (2001)
COMMENT
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border of

AUTHORS	Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.
TITLE	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL	Unpublished (2001)
COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu This is single pass sequence recovered from the left border of TDNA. Class: TDNA tagged.
FEATURES	location/Qualifiers 1..384 /organism="Arabidopsis thaliana" /strain="Columbia 0" /db_xref="taxon:3702" /clone_lib="SALK_010743" /note="PCR was performed on Arabidopsis thaliana lines each of which contains one or more TDNA insertion elements. The resultant fragment for each line was directly sequenced to determine the genomic sequence at the site of insertion. Details of the protocols used can be found at http://signal.salk.edu/tdna_protocols.html"
BASE COUNT	87 a 127 c 51 g 119 t
ORIGIN	Query Match 4.2%; Score 42.8; DB 17; Length 384; Best Local Similarity 75.7%; Pred. No. 0.046; Matches 53; Conservative 0; Mismatches 17; Indels 0; Gaps 0;
QY	1 CTTACTATAGGCGACGGTGTCTGCAGCGCCGGCTGTATGAAGTGCGACCTCACTG 60 DB 106 CTCACATATAGGCGACGGTGTCTGCAGCGCCGGCTGTATGAAGTGCGACCTCAATG 47 QY 61 GATGCATATA 70 Db 46 AAGTAAATA 37
RESULT 6	BH634101/c 445 bp DNA linear GSS 01-FEB-2002
LOCUS	BH634101
DEFINITION	SALK_044702 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_044702, DNA sequence.
ACCESSION	BH634101
VERSION	BH634101.1 GI:18469464
KEYWORDS	GSS.
SOURCE	thale cress. ORGANISM Arabidopsis thaliana Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsi
REFERENCE	1 (bases 1 to 445) Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R. A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome Unpublished (2001) Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu This is single pass sequence recovered from the left border of

QY 64 GCATATAGTC 74
 Db 199 GAATAACTCC 189

RESULT 9
 BH214178/c
 LOCUS
 DEFINITION SALK_010362 Arabidopsis thaliana TDNA insertion lines Arabidopsis
 thaliana genomic clone SALK_010362, DNA sequence.

ACCESSION
 VERSION BH214178.1 GI:16396703
 KEYWORDS
 SOURCE
 ORGANISM Arabidopsis thaliana
 thale cress.

REFERENCE
 AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
 ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.,
 Zimmerman,J. and Ecker,J.R. 1999. Arabidopsis thaliana
 A Sequence-indexed Library of Insertion Mutations in the
 Arabidopsis Genome
 Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (SIGNAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of
 TDNA. This sequence lies within an exon of At5g22340.
 Class: TDNA tagged.

FEATURES
 source
 Location/Qualifiers
 1..500
 /organism="Arabidopsis thaliana"
 /strain="Columbia 0"
 /db_xref="taxon:3702"
 /clone="SALK_010362"
 /note="PCR was performed on Arabidopsis thaliana lines
 each of which contains one or more TDNA insertion
 elements. The resultant fragment for each line was
 directly sequenced to determine the genomic sequence at
 the site of insertion. Details of the protocols used can
 be found at http://signal.salk.edu/tdna_protocols.html"

BASE COUNT 54 a 260 c 41 g 145 t

Query Match 4.1%; Score 42.2; DB 17; Length 500;
 Best Local Similarity 79.4%; Pred. No. 0.08;
 Matches 50; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 1 CTCTACTAGGACCGTGTCTGCGCGCGCGGCTGATGAAGTGGGAACCTCAGTG 60
 Db 141 CTCTACTAGGACCGGTTGCTGCGCGCGCGGCTGATGAAGTGGGAGTCCGCG 82

QY 61 GAT 63
 Db 81 GAT 79

RESULT 10
 CNS00LGV
 LOCUS
 DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
 BACR30M12 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.

ACCESSION
 VERSION AL068402
 KEYWORDS
 SOURCE
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see http://www.fruitfly.org The BDGP Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mammot in Pletier de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain v2; cn bw sp, the same strain used for the BDGP's
 P1 and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 Location/Qualifiers
 1..1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR30M12"
 /clone_lib="RPCI-98"
 /note="end : T7"

BASE COUNT 337 a 227 c 208 g 266 t 63 others

Query Match 4.1%; Score 42.2; DB 17; Length 1101;
 Best Local Similarity 44.8%; Pred. No. 0.11;
 Matches 90; Conservative 16; Mismatches 95; Indels 0; Gaps 0;

QY 260 TAGTAAGCAGGCCCGCCAGCATTCAAATTCGATGTAATTTGTCTACTTCATTTGGTTCA 319
 Db 856 TAGATAGCAAAATATACCATTCGCATAAATATATATGTCTTCTTAAAGCCAGTGAA 915

QY 320 GATTTGAACGATGTCAGTTTCAATCCACATCCACAGATCCGACAGAGATTCATCAAA 379
 Db 916 CACTTTAGCCATTTGAAGAGGCGATTTAAAGGGATGCGGTCCAGGAATTCGGAAGTGW 975

QY 380 TTCACACATAAGCAGCAGCAGAGTAAACCATCAATCAATCAATCAATCAATCAATTT 439
 Db 976 AGCAGACAGACAGCAAAARCAAGAGTCGAGGAGCAGAVCAAGCAAGCAAGCAAGCAAG 1035

QY 440 TAGATCTTTTGAAGAGGACAA 460
 Db 1036 AWTAACTTTGGCAAAAGMAAA 1056

RESULT 11
 BH168812/c
 LOCUS
 DEFINITION SALK_000216 Arabidopsis thaliana TDNA insertion lines Arabidopsis
 thaliana genomic clone SALK_000216, DNA sequence.

ACCESSION
 VERSION BH168812
 KEYWORDS
 SOURCE
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 129)
 Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab
 ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P.

Query Match	4.1%	Score 42;	DB 17;	Length 220;	
Best Local Similarity	73.0%;	Pred. No. 0.067;			
Matches 54;	Conservative 0;	Mismatches 20;	Indels 0;	Gaps 0	
QY	1	CTTACTATAGGCGACCGCTGGTCGACGGCGCGGGCTGTATGATGAGGTGGGAAACCTCACTG	60		
Db	214	CTCACTATAGGCGACCGCTGGTCGACGGCGCGGGCTGCAATTCAAGTCCTAACACACAATT	155		
QY	61	GATGCATATATGCG	74		
Db	154	TTTTTAATTAGTGC	141		

RESULT 13	
BH212774/C	
LOCUS	
DEFINITION	
ACCESSION	BH212774
VERSION	BH212774.1
KEYWORDS	GSS.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (bases 1 to 452)

```

TITLE      A Sequence-Indexed Library of Insertion Mutations in the
JOURNAL    Arabidopsis Genome
COMMENT     Unpublished (2001)
           Contact: Joseph R. Ecker
           Salk Institute Genomic Analysis Laboratory (SIGNAL)
           The Salk Institute for Biological Studies
           10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
           Tel: 858 453 4100 x1752
           Fax: 858 558 6379
           Email: ecker@salk.edu

           This is single pass sequence recovered from the left border of
           TDNA.

Class:     TDNA tagged.
           Location/Qualifiers
             1. .452
             /organism="Arabidopsis thaliana"
             /strain="Columbia 0"
             /db_xref="taxon:3702"
             /clone="SALK_008030"
             /clone_lib="Arabidopsis thaliana TDNA insertion lines"
             /note="PCR was performed on Arabidopsis thaliana lines
             each of which contains one or more TDNA insertion
             elements. The resultant fragment for each line was
             directly sequenced to determine the genomic sequence at
             the site of insertion. Details of the protocols used can
             be found at http://signal.salk.edu/tdna\_protocols.html"
           114 a 113 c 111 g 93 t 21 others
BASE COUNT

```

[illegible]

Db 192 GACGT

Search completed: December 28, 2002, 00:00:39
Job time : 1973 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: December 27, 2002, 23:39:28 ; Search time 2660 Seconds
(without alignments)
11269.132 Million cell updates/sec

Title: US-09-945-376-3
Perfect score: 1030
Sequence: 1 cttactataggcagcgtg.....tgatatatacaagggtctgg 1030

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg_hum.*
- 31: em.htg_inv.*
- 32: em.htg_other.*
- 33: em.htg_mus.*
- 34: em.htg_pin.*
- 35: em.htg_rod.*
- 36: em.htg_man.*
- 37: em.htg_vrt.*
- 38: em.sy.*
- 39: em.htgo_hum.*
- 40: em.htgo_mus.*
- 41: em.htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	49.8	4.8	125020	9	AF429315	AF429315 Homo sapi
2	44.8	4.3	125020	9	AF429315	AF429315 Homo sapi
3	44.4	4.3	216878	2	AL773534	AL773534 Mus muscu
4	42.8	4.2	685	8	AF457660	AF457660 Castanea
5	42.4	4.1	1423	9	AF039526	AF039526 Homo sapi
6	42.4	4.1	6078	6	ARI24194	ARI24194 Sequence
7	41.8	4.1	92633	2	PFMAL4P1_3	Continuation (4 of
C 8	41.8	4.1	110000	2	PFMAL4P1_2	Continuation (3 of
C 9	41	4.0	4260	5	AY124482	AY124482 Danio rer
C 10	41	4.0	11957	9	AF051769	AF051769 Homo sapi
C 11	40.8	4.0	1094	8	AF075270	AF075270 Hordeum v
C 12	40.8	4.0	166703	9	AL139378	AL139378 Human DNA
13	40.6	3.9	156325	9	HSL34N8	AL031655 Human DNA
14	40.4	3.9	598	6	AX320057	AX320057 Sequence
15	40.4	3.9	1463	3	AF153014	AF153014 Trichomon
C 16	40.4	3.9	2611	6	AX380786	AX380786 Sequence
C 17	40.2	3.9	2463	9	AB021922	AB021922 Homo sapi
C 18	40.2	3.9	220031	9	AC025822	AC025822 Homo sapi
19	40	3.9	498	10	RNU53907	U53907 Rattus norv
20	40	3.9	951	3	AF233737	AF233737 Agrotis i
21	40	3.9	985	6	AX253432	AX253432 Sequence
22	40	3.9	1499	6	AX253428	AX253428 Sequence
23	40	3.9	1958	10	AF114032	AF114032 Mus muscu
24	39.8	3.9	682	10	MMU289605	AJ289605 Mus muscu
25	39.8	3.9	2385	6	AX041981	AX041981 Sequence
C 26	39.8	3.9	34980	6	AX444560	AX444560 Sequence
27	39.6	3.8	7218	6	I66494	I66494 Sequence 14
28	39.4	3.8	640	10	MMU63899	U63899 Mus musculu
29	39.4	3.8	1135	9	AF074397	AF074397 Homo sapi
C 30	39.4	3.8	2492	6	AX098431	AX098431 Sequence
C 31	39.4	3.8	2867	6	E26792	E26792 Novel prote
C 32	39.4	3.8	4185	1	AF237414	AF237414 Ehrlichia
33	39	3.8	516	6	AX320056	AX320056 Sequence
34	39	3.8	615	10	AF077743	AF077743 Mus muscu
35	39	3.8	2184	6	AX253431	AX253431 Sequence
36	39	3.8	2785	10	AF162890S1	AF162890 Mus muscu
37	38.8	3.8	104526	9	AC087887	AC087887 Homo sapi
C 38	38.8	3.8	149440	2	AC024319	AC024319 Homo sapi
C 39	38.8	3.8	196049	2	AC105841	AC105841 Rattus no
C 40	38.8	3.8	299350	1	SME591786	AL591786 Sinorhizo
41	38.8	3.8	329100	1	SME591787	AL591787 Sinorhizo
42	38.6	3.7	1157	8	AF457661	AF457661 Castanea
C 43	38.6	3.7	5031	6	AX085162	AX085162 Sequence
C 44	38.6	3.7	5031	6	AX085359	AX085359 Sequence
45	38.6	3.7	180968	2	AC078996	AC078996 Mus muscu

ALIGNMENTS

RESULT 1
AF429315/c
LOCUS
DEFINITION Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
ACCESSION AF429315
VERSION AF429315.1 GI:17646244
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 125020)
AUTHORS Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.

TITLE A repeat expansion in the gene encoding juncctophilin-3 is associated with Huntington disease-like 2

JOURNAL Nat. Genet. 29 (4), 377-378 (2001)

MEDLINE 21583737

PUBMED 11694876

REFERENCE 2 (bases 1 to 125020)

AUTHORS Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.

TITLE Direct Submission

JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA

FEATURES

source 1..125020

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/map="16q24.3; between D16S520 and WI-12410"

/note="isolated from a patient with Huntington's Disease-Like 2 (HDL2)"

repeat_region complement(35581..35746)

/rpt_unit=ctg

/rpt_type=tandem

/gene="JPH3"

/note="JPH3"

mRNA complement(<36507..>36887)

/gene="JPH3"

CDS complement(<36507..36887)

/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"

/codon_start=1

/product="juncctophilin 3"

/protein_id="AAL40941.1"

/db_xref="GI:17646245"

/translation="MSSGGRFNFDGSGYCGWEDGKAHGHCVCTGPKGQGEVTSWS HGFELGVYTWPSNTYQGTWAGQKRGHIGLESKGKWKYKGETHGFKRGYGVRECA NGAKYEGTWSNGLQDGYGTETYSDG"

BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others

ORIGIN

Query Match 4.3%; Score 49.8; DB 9; Length 125020;

Best Local Similarity 9.38; Pred. No. 0.007;

Matches 41; Conservative 213; Mismatches 185; Indels 1; Gaps 1;

QY 525 AGCCTGAACCGCTCGTATTGTCAGTTGGGGTTGGTGGCCAGCCGATGATGCCATCC 584

Db 17666 WSSYSTSMGYSTSCCKYKCSWSSMYKCKTSKSRKRSYIYWGGRKAKKYYCAGRR 17607

QY 585 GAACCTGCTCCAACTCGATTTCCGTTTCAGAGCAGTACTGCAATTTGCCATTTGTCG 644

Db 17606 RMSYWKCCAKWMSYCCWSYCMYTYYSKCTYKSTCYKRGGYWGSKTCYSAGGKRS 17547

QY 645 CTCGACGAGAGACTCGTACTGTC-CGCGCGGAACCGACCTCCCTCTGCTGTCGCCA 703

Db 17546 MYTCMKRSSSSMSMARSSMCWGAGYRRKRSKAGWAGWRSRSGKRGSTGKMRACS 17487

QY 704 ACTGCCAACAGACCTGGCATGCCCTGGTGGGCTATCTAATTTCTCCATCAAGAAATC 763

Db 17486 KKTGSGTGRSMKKKGYSKYSRGMKKKCTCYCMWKYKRYKRTSMCWYTHKSWGKR 17427

QY 764 ACTCCCATTAATTCGCTGTGGCGGACCTCGTGCCTCGAATTTCCAGGCGCACGG 823

Db 17426 YKRCMKKKGCTGYRMSKSKSGYKSMRMRGSSYSTSCWKGYSWMKCMKYSYKK 17367

QY 824 TGCAGGCACACACCCCATCAAAATACCGCCTCTTTGGCGGACCGAATCTGCACCGAG 883

Db 17366 RSMRSGMSKMGWYAGGCGCTSSSMWSTKRKRSCYSYKSKYKGRKMGKMGKGRSK 17307

QY 884 GCGCAATTAATTCGAGCCTGATCGAGGTGGTTGGATCGCGCAGGAAACCTTGAATGCG 943

Db 17306 YWSSMKMKRSSWSKCYSTKYSGRKRSKGRSTKSKAKSSMRAGSKCTYGSYSWN 17247

QY 944 CGGCATCAGTAGCAGCGGCA 963

Db 17246 NRRNRMGKTGCNYMYRRSR 17227

RESULT 2

LOCUS AF429315 125020 bp DNA linear PRI 18-JAN-2002

DEFINITION Homo sapiens juncctophilin 3 (JPH3) gene, partial cds.

ACCESSION AF429315

VERSION AF429315.1 GI:17646244

KEYWORDS Homo sapiens.

SOURCE Homo sapiens.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 125020)

AUTHORS Holmes, S.E., O'Hearn, E., Rosenblatt, A., Callahan, C., Hwang, H.S., Ingersoll-Ashworth, R.G., Fleisher, A., Stevanin, G., Brice, A., Potter, N.T., Ross, C.A. and Margolis, R.L.

TITLE A repeat expansion in the gene encoding juncctophilin-3 is associated with Huntington disease-like 2

JOURNAL Nat. Genet. 29 (4), 377-378 (2001)

MEDLINE 21583737

PUBMED 11694876

REFERENCE 2 (bases 1 to 125020)

AUTHORS Holmes, S.E., Ingersoll-Ashworth, R.G., Ross, C.A. and Margolis, R.L.

TITLE Direct Submission

JOURNAL Submitted (05-OCT-2001) Psychiatry, Johns Hopkins Medical Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA

FEATURES

source 1..125020

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/map="16q24.3; between D16S520 and WI-12410"

/note="isolated from a patient with Huntington's Disease-Like 2 (HDL2)"

repeat_region complement(35581..35746)

/rpt_unit=ctg

/rpt_type=tandem

/rpt_unit=ctg

gene complement(<36507..>36887)

/gene="JPH3"

mRNA complement(<36507..>36887)

/gene="JPH3"

CDS complement(<36507..36887)

/gene="JPH3"

/note="component of the junctional complex between plasma membrane and endoplasmic reticulum"

/codon_start=1

/product="juncctophilin 3"

/protein_id="AAL40941.1"

/db_xref="GI:17646245"

/translation="MSSGGRFNFDGSGYCGWEDGKAHGHCVCTGPKGQGEVTSWS HGFELGVYTWPSNTYQGTWAGQKRGHIGLESKGKWKYKGETHGFKRGYGVRECA NGAKYEGTWSNGLQDGYGTETYSDG"

BASE COUNT 29056 a 32731 c 30696 g 28283 t 4254 others

ORIGIN

Query Match 4.3%; Score 44.8; DB 9; Length 125020;

Best Local Similarity 10.1%; Pred. No. 0.22;

Matches 53; Conservative 260; Mismatches 207; Indels 4; Gaps 1;

QY 470 CAGACGACAAAATTCGGCATCAACACCTTTTCGCCACAAAGTAATGTTGGGTAGGCC 529

Db 50579 CTGRWTGCBWSSKSHSSCHRRRTMDYRMRYTYCCCTATYMMBYHYCYDBCWSGVWGSY 50638

QY 530 TGAACCGCTCGCTATTGCAGTTGGGTGGTGGCAGCCCGATGATCTCCATCCGAACT 589

Db 50639 GGWRGMMKMBDBMBBSMGMKSGMDGTDKDKYWCYSSMDVVBWRWBDTWSGSBSHYMR 50698

Db 51059 RSGRMSAYCSCRKCKSKRCSCGCMRCSWYCWYKCMRYGSGWRMYKC 51102

```
1. 003
source
/organism="Castanea dentata"
/db xref="taxon:134033"
```



```
|||||
Db 107315 CTTACTATAGGCACGCGTGTCTGCACGCCGCGTGGTGAATTTTTTTTTTTTTTTT 107256
QY 61 GATGCATATACTGCTGAGAGATACATCACAAAT 93
|||||
Db 107255 TTTCATATTTTAAAGGCTCATATAATAAT 107223
|||||

RESULT 9
AY124482/c 4260 bp DNA linear VRT 31-JUL-2002
LOCUS
DEFINITION Danio rerio myogenin gene, exons 1, 2 and 3 and complete cds.
ACCESSION AY124482
VERSION AY124482.1 GI:22035404
KEYWORDS
SOURCE
ORGANISM zebrafish.
Danio rerio
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 4260)
AUTHORS Du,S., Gao,J. and Anyangwe,V.
TITLE Muscle-specific expression of myogenin in zebrafish embryos is
controlled by multiple regulatory elements in the promoter
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4260)
AUTHORS Du,S., Gao,J. and Anyangwe,V.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2002) Center of Marine Biotechnology, University
of Maryland Biotechnology Institute, 701 E. Pratt St, Baltimore, MD
21202, USA
FEATURES
source Location/Qualifiers
1..4260
/organism="Danio rerio"
/db_xref="taxon:7955"
1..651
promoter join(652..1309,2288..2393,3002..3599)
mRNA /product="myogenin"
552..11309
exon /number=1
join(764..1309,2288..2393,3002..3120)
CDS /codon_start=1
/product="myogenin"
/db_xref="GI:22035405"
/protein_id="AA822616.1"
/translating="MELFETNPYFFNDQRFYEGADNFFQSRINGFEQAGYQDRNSMM
GLGDGRMLTTTVGLEDKPSPSSSLGLMSPHOQHCPCQCLPWACKYCKRKSVTMD
RKAATLRKRRLLKYNFAEALKRSTLMNPORLPKVEILRSAYOYIERLOALVSSL
NOQHEQGNLHYRATAAPHTGVSSSDGSGTCCSPWSSASDHCVPAYSSAHD
LLNDSSEQNLSLSIVDSITGTATPVAYSDISK"
2288..2393
exon /number=2
3002..3599
exon /number=3
BASE COUNT 1361 a 797 c 731 g 1371 t
ORIGIN
Query Match 4.0%; Score 41; DB 5; Length 4260;
Best Local Similarity 89.8%; Pred. No. 1.5;
Matches 44; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 TTACTATAGGCACGCGTGTCTGCACGCCGCGTGGTGAATGAAGTGGG 50
|||||
Db 4242 TTACTATAGGCACGCGTGTCTGCACGCCGCGTGGTGAATGAAGTGGG 4194
|||||

RESULT 10
AF051769/c 11957 bp DNA linear PRI 31-OCT-1999
LOCUS
DEFINITION Homo sapiens hyaluronidase-like pseudogene 1 (HYALP1), partial
sequence.
ACCESSION AF051769
VERSION AF051769.1 GI:6164589
|
```

5' UTR /gene="HVST1"
1025..1089 /product="high affinity sulfate transporter"
CDS /gene="HVST1"
1090..>1094 /codon_start=1
/product="high affinity sulfate transporter"
/protein_id="AAC42664.1"
/db_xref="GI:3643824"
/translation="Mp"
BASE COUNT 286 a 261 c 243 g 304 t
ORIGIN

Query Match 4.0%; Score 40.8; DB 8; Length 1094;
Best Local Similarity 95.5%; Pred. No. 1.3;
Matches 42; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 CTTACTATAGGCACGCGTGTGCGACGCCCGGGCTGGTATGAA 44
|||||
Db 63 CTTACTATAGGCACGCGTGTGCGACGCCCGGGCTGCTGCA 106
|||||

RESULT 12
AL139378/c
LOCUS AL139378 166703 bp DNA linear PRI 02-DEC-2000
DEFINITION Human DNA sequence from clone RP11-271B5 on chromosome 13 Contains
a gene for a protein similar to ribosomal protein S7, the FGF9
(fibroblast growth factor 9 (glia-activating factor)) gene, ESTs,
STSs, GSSs and CpG islands, complete sequence.
AL139378
AL139378.15 GI:8247510
VERSION HTG; CpG island; FGF9; growth factor; ribosomal protein.
KEYWORDS human.
SOURCE Homo sapiens

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 166703)
Blakey, S.
Direct Submission
Submitted (18-JUL-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerquest@sanger.ac.uk

On Jun 4, 2000 this sequence version replaced gi:8246895.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em', EMBL; Sw', SWISSPROT; Tr', TREMBL; Wp', WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C-elegans/wormpep this sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr13
RP11-271B5 is from the library RPCI-11.1 constructed at the Roswell
Park Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/
VECTOR: pBACE3.6
This sequence is the entire insert of clone RP11-271B5 The true
left end of clone RP11-282A24 is at 82919 in this sequence. The
true right end of clone RP11-110K8 is at 72244 in this sequence.
Location/Qualifiers

source
1..166703
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-271B5"
/clone_lib="RPCI-11.1"
27..485
/note="match: GSS: Em:B83261"
463..580
/note="MERSA repeat: matches 17..138 of consensus"
1191..1400
/note="MER3 repeat: matches 1..209 of consensus"
1669..1922
/note="AluJb repeat: matches 1..294 of consensus"
2320..2372
/note="L2 repeat: matches 2700..2750 of consensus"
2621..2932
/note="MLTIC repeat: matches 147..466 of consensus"
2933..3198
/note="L1PA16 repeat: matches 5853..6157 of consensus"
3250..3410
/note="MLTIC repeat: matches 1..151 of consensus"
4184..4319
/note="L2 repeat: matches 2609..2748 of consensus"
4703..4947
/note="L2 repeat: matches 1807..2082 of consensus"
5016..5305
/note="AluX repeat: matches 1..288 of consensus"
5493..5804
/note="Alu repeat: matches 1..311 of consensus"
6576..6869
/note="AluSc repeat: matches 1..289 of consensus"
7077..7157
/note="MSTC repeat: matches 291..359 of consensus"
7324..7630
/note="Alu repeat: matches 1..304 of consensus"
7766..7889
/note="FLAM_C repeat: matches 11..132 of consensus"
8776..9085
/note="AluX repeat: matches 5..312 of consensus"
9516..9818
/note="Alu repeat: matches 1..303 of consensus"
10023..10311
/note="Alu repeat: matches 1..293 of consensus"
complement(10938..11116)
/note="match: GSS: Em:AQ384546"
complement(10938..11050)
/note="match: GSS: Em:AQ219188"
complement(10953..11555)
/gene="BA271B5.1"
/note="match: CDNAS: Em:M77233
match: ESTs: Em:AA105661 Em:A1462024
match: proteins: Tr:Q9ZNS1 Sw:Q10101 Sw:O43105 Tr:Q9VA91
Tr:Q23312 Sw:P02362 Sw:P23821 Sw:P48155 Sw:P33514
Tr:Q29190 Tr:O13167 Sw:P50894"
/codon_start=1
/evidence=not_experimental
/product="BA271B5.1 (similar to ribosomal protein S7)"
/protein_id="CAC17691.1"
/db_xref="GI:11544936"
/translation="MFSSSAKIVKPNDEKDPDEFSGISQALLEMNSDLKAOLRELN
ITAAKTEILGGGRKAIIFVPIQLKSFQKTOVELVRELEKFKSGKHVVIVQRILP
KPKRSRTKKQKRPRSRHTLTAVDAILEDLVFPSEIVGKRIRVKLDGSRILKRVHLDK
AQNNVSHKVTSGVYKLTGKDVNFEFFELQTKMTT"
complement(10958..11260)
/gene="BA271B5.1"
/note="match: GSS: Em:AQ887379"
10995..11131
/note="match: GSS: Em:AQ476700"
10997..11430
/note="match: GSS: Em:AQ593042"
misc_feature
misc_feature
misc_feature

misc_feature 11256..11498 /note="match: GSS: Em:AQ264373"
misc_feature 11281..11566 /note="match: GSS: Em:AQ554409"
repeat_region 11672..11854 /note="MER5A repeat: matches 3..188 of consensus"
repeat_region 11884..12154 /note="L1M4 repeat: matches 2891..3182 of consensus"
repeat_region 12941..13266 /note="MER2 repeat: matches 1..345 of consensus"
repeat_region 13779..13989 /note="L1M4C repeat: matches 1511..1736 of consensus"
repeat_region 14084..14294 /note="L1M4C repeat: matches 1260..1464 of consensus"
repeat_region 15315..15827 /note="AluY repeat: matches 1..303 of consensus"
repeat_region 16476..16517 /note="L14 copies 3 mer tta 78% conserved"
repeat_region 16651..16946 /note="AluSx repeat: matches 1..296 of consensus"
repeat_region 17442..17709 /note="AluSx repeat: matches 6..269 of consensus"
repeat_region 18944..19037 /note="MIR repeat: matches 49..143 of consensus"
misc_feature 19023..19447 /note="match: GSS: Em:AQ630744"
misc_feature 19036..19496 /note="match: GSS: Em:AQ630723"
repeat_region 19722..19854 /note="MER5B repeat: matches 1..138 of consensus"
repeat_region 19863..20033 /note="MER5A repeat: matches 1..181 of consensus"
repeat_region 20255..20388 /note="L2 repeat: matches 2612..2750 of consensus"
repeat_region 23464..23487 /note="L2 copies 2 mer tg 100% conserved"
repeat_region 23500..23813 /note="AluSx repeat: matches 1..312 of consensus"
repeat_region 25990..26066 /note="L2 repeat: matches 2662..2739 of consensus"
repeat_region 26385..26452 /note="L2 copies 34 mer 100% conserved"
repeat_region 26461..26919 /note="MLN2CB repeat: matches 1..494 of consensus"
misc_feature 26705..27376 /note="match: GSS: Em:AQ379984"
repeat_region 26925..27220 /note="L2 repeat: matches 1..295 of consensus"
repeat_region 27792..27888 /note="L2 repeat: matches 2597..2697 of consensus"
misc_feature 28519..28685 /note="match: GSS: Em:AQ559377"
repeat_region 29089..29230 /note="MER5B repeat: matches 6..177 of consensus"
repeat_region 29406..29491 /note="MER5A repeat: matches 74..159 of consensus"
repeat_region 29902..30303 /note="L1PA5 repeat: matches 5740..6145 of consensus"
misc_feature 30726..31152 /note="match: GSS: Em:AQ377301"
repeat_region 31396..31605 /note="MLR2G repeat: matches 228..386 of consensus"
repeat_region 31899..32195 /note="AluY repeat: matches 1..297 of consensus"
repeat_region 32459..32670 /note="AluY repeat: matches 1..303 of consensus"
repeat_region 32786..32818 /note="MLR2FA repeat: matches 96..128 of consensus"
repeat_region 32843..33060 /note="MLR2E repeat: matches 121..349 of consensus"
repeat_region 33068..33143

repeat_region 33069..33140 /note="19 copies 4 mer ctct 65% conserved"
repeat_region 33401..33472 /note="36 copies 2 mer tc 65% conserved"
repeat_region 33402..33473 /note="18 copies 4 mer atat 81% conserved"
Query Match 4.0%; Score 40.8; DB 9; Length 166703;
Best Local Similarity 54.7%; Pred.No.3.5; Mismatches 67; Indels 0; Gaps 0;
Matches 81; Conservative 0;
QY 303 TACTTCATTGTGTTGAGAGATGTCAGTTTCAATCCACATGCAACAGATCG 362
Db 48869 TACAATATTTAAATATTTTAAAGCTATAATAATTAGACAGTGTCTATGAATTAT 48810
QY 363 CAGAAGAAATCATCCAAATTCACACATAAAGCAGCAACAGAAAGTTAAACCATCAACA 422
Db 48809 TTGTATTAGCTATGAATAAACAGATAAAGCAATACAGAAATAGAAATTACATAGATCCA 48750
QY 423 ATTAACGAGTTGCATTTTAGATCTTTTA 450
Db 48749 AATACATATTAGAAATTTTATATATGATA 48722
RESULT 13
HS134N8
LOCUS Human DNA sequence from clone RP1-134N8 on chromosome 20p12.
DEFINITION Contains STSS, GSSs and a CpG island, complete sequence.
ACCESSION AL031655
VERSION AL031655.8 GI:12545137
KEYWORDS HTG; CpG island.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 156325)
Ramsay H.
Direct Submission
Submitted (05-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
Requests: clonerequest@sanger.ac.uk
On Jan 26, 2001 this sequence version replaced gi:5360979.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em.: EMBL; Sv.: SWISSPROT; Tr.: TrEMBL; Wp.: WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C.elegans/wormpep
This sequence
was generated from part of bacterial clone contigs of human
chromosome 20, constructed by the Sanger Centre Chromosome 20
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr20
RP1-134N8 is from the library RPC1-1 constructed by the group of
Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pCYPAC2
This sequence is the entire insert of clone RP1-134N8 The true left
end of clone RP5-1007P8 is at 84660 in this sequence. The true
right end of clone RP5-913G4 is at 91640 in this sequence. This
sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest.
Location/Qualifiers

FEATURES

1 (bases 1 to 1463)
Crouch, M.V. and Alderete, J.F.
Trichomonas vaginalis has two fibronectin-like iron-regulated genes
Arch. Med. Res. 32 (2), 102-107 (2001)
21242390
2 (bases 1 to 1463)
Crouch, M.-L. and Alderete, J.F.
Direct Submission
Submitted (19-MAY-1999) Microbiology, UTHSCSA, 7703 Floyd Curl Dr,
San Antonio, TX 78284-7758, USA

1. .1463

Sequence	519 a	242 C	233 g	469 t	
Sequence	DHLVLRNRGSRPPECTETPL"				
Sequence	VALFKLVETKYEKHEQYDGPDKETIYDVVKKKVFNQDPKIQEYCKTLEVGKKRYRVC				
Identity	3.9%; Score 40.4; DB 3; Length 1463;				
Similarity	97.6%; Pred. No. 1.9;				
Conservative	0; Mismatches 1; Indels 0 Gaps 0;				

1 Similarity 97.6%; Pred. No. 1.9;
41; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
ACTATATAGGCACGCGTGGTCTGACGGCCGGGTGGTATGAA 44
|||||
ACTATATAGGCACGCGTGGTCTGACGGCCGGGTGGTCTGAA 52

leted: December 28, 2002, 01:04:17
3375 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 27, 2002, 20:39:18 ; Search time 55 Seconds
(without alignments)
5743.219 Million cell updates/sec

Title: US-09-945-376-3
Perfect score: 1030
Sequence: 1 ctctactataggccagcgctg.....tgtatatatacaaggtctgg 1030

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2.6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2.6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2.6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2.6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2.6/ptodata/1/ina/PTUS_COMB.seq:*
6: /cgn2.6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	42.4	4.1	6078	4 US-09-173-914-1	Sequence 1, Appli
2	39.8	3.9	3718	4 US-09-424-283-6	Sequence 6, Appli
3	39.6	3.8	7218	1 US-08-232-463-14	Sequence 14, Appl
4	39	3.8	341	4 US-09-323-195A-1	Sequence 1, Appli
5	39	3.8	4526	4 US-09-424-283-7	Sequence 7, Appli
6	38.4	3.7	921	4 US-09-377-648-4	Sequence 4, Appli
7	38.4	3.7	2791	4 US-09-570-367C-1	Sequence 1, Appli
8	38	3.7	336	4 US-09-276-599-13	Sequence 13, Appl
9	38	3.7	763	4 US-09-276-599-14	Sequence 14, Appl
10	37.8	3.7	565	4 US-09-323-195A-5	Sequence 5, Appli
11	37.8	3.7	1924	4 US-09-424-283-5	Sequence 5, Appli
12	37.4	3.6	48	4 US-08-913-014A-18	Sequence 18, Appl
13	37.4	3.6	2791	4 US-09-570-367C-1	Sequence 1, Appli
14	37.2	3.6	1478	4 US-09-545-814-28	Sequence 28, Appl
15	37.2	3.6	1478	4 US-09-545-814-30	Sequence 30, Appl
16	37.2	3.6	1677	4 US-09-545-814-13	Sequence 13, Appl
17	37.2	3.6	1677	4 US-09-545-814-15	Sequence 15, Appl
18	37.2	3.6	1749	4 US-09-545-814-4	Sequence 4, Appli
19	37.2	3.6	1749	4 US-09-545-814-6	Sequence 6, Appli
20	37.2	3.6	1919	4 US-09-545-814-31	Sequence 31, Appl
21	37.2	3.6	1919	4 US-09-545-814-33	Sequence 33, Appl
22	37.2	3.6	2610	4 US-09-545-814-1	Sequence 1, Appli
23	37.2	3.6	2610	4 US-09-545-814-3	Sequence 3, Appli
24	37	3.6	340	4 US-09-323-195A-3	Sequence 3, Appli
25	36	3.5	255	4 US-09-257-583-5	Sequence 5, Appli
26	36	3.5	342	4 US-09-323-195A-6	Sequence 6, Appli
27	36	3.5	555	4 US-08-905-223-37	Sequence 37, Appl

28	36	3.5	555	4	US-09-247-155-37	Sequence 37, Appl	
29	35	3.4	189	4	US-09-134-001C-1688	Sequence 1688, Ap	
30	34.8	3.4	47	2	US-08-582-562A-53	Sequence 53, Appl	
31	34.8	3.4	47	2	US-08-778-494B-53	Sequence 53, Appl	
32	34.8	3.4	1881	2	US-08-596-300A-4	Sequence 4, Appli	
33	34.8	3.4	3041	2	US-08-596-300A-5	Sequence 5, Appli	
34	34.8	3.4	3718	2	US-08-596-300A-6	Sequence 6, Appli	
35	34.8	3.4	3796	2	US-08-596-300A-3	Sequence 3, Appli	
36	33	3.2	1440	4	US-09-134-001C-1291	Sequence 1291, Ap	
37	32.2	3.1	340	4	US-09-323-195A-4	Sequence 4, Appli	
38	32.2	3.1	396	4	US-08-887-534A-21	Sequence 21, Appl	
c	39	3.2	3.1	500	3	US-08-755-587-37	Sequence 37, Appl
40	31.8	3.1	1218	3	US-08-976-122-1	Sequence 1, Appli	
41	31.4	3.0	3064	4	US-09-600-776-5	Sequence 5, Appli	
c	42	31.4	3.0	246240	2	US-08-724-394A-20	Sequence 20, Appl
c	43	31.4	3.0	246240	2	US-08-724-394A-21	Sequence 21, Appl
c	44	31.4	3.0	246240	2	US-08-724-394A-22	Sequence 22, Appl
c	45	31.2	3.0	6202	1	US-08-484-101B-41	Sequence 41, Appl

ALIGNMENTS

RESULT 1
US-09-173-914-1
; Sequence 1, Application US/09173914
; Patent No. 6171857
; GENERAL INFORMATION:
; APPLICANT: Hendrickson, Eric
; TITLE OF INVENTION: A No. 6171857el Leucine Zipper, KARP-1 and
; FILE REFERENCE: B0877/7017/HK
; CURRENT APPLICATION NUMBER: US/09/173,914
; CURRENT FILING DATE: 1998-10-16
; EARLIER APPLICATION NUMBER: 60/064,557
; EARLIER FILING DATE: 1997-10-17
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 6078
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (178)...(178)
; NAME/KEY: unsure
; LOCATION: (230)...(230)
; NAME/KEY: unsure
; LOCATION: (232)...(232)
; NAME/KEY: unsure
; LOCATION: (234)...(234)
; NAME/KEY: unsure
; LOCATION: (453)...(453)
; NAME/KEY: unsure
; LOCATION: (473)...(473)
; NAME/KEY: unsure
; LOCATION: (610)...(610)
; NAME/KEY: unsure
; LOCATION: (612)...(612)
; NAME/KEY: unsure
; LOCATION: (2175)...(2175)
; NAME/KEY: unsure
; LOCATION: (1014)...(1014)
US-09-173-914-1

Query Match 4.1%; Score 42.4; DB 4; Length 6078;
Best Local Similarity 97.7%; Pred. No. 0, 0041;
Matches 43; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 CTTACTATAGGCACGCGTGTCTGACGGCCCGGGCTGGTATGAA 44
|||||
Db 9 CTTACTATAGGCACGCGTGTCTGACGGCCCGGGCTGGTCTGAA 52
|||||

```
RESULT 2
US-09-424-283-6
; Sequence 6, Application US/09424283
; Patent No. 6437219
; GENERAL INFORMATION:
; APPLICANT: Grimes, et al.
; TITLE OF INVENTION: Sucrose binding proteins
; FILE REFERENCE: 4630-50206
; CURRENT APPLICATION NUMBER: US/09/424,283
; CURRENT FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: PCT/US98/10465
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: US 60/047,568
; PRIOR FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 3718
; TYPE: DNA
; ORGANISM: Glycine max
US-09-424-283-6

Query Match          3.9%; Score 39.8; DB 4; Length 3718;
Best Local Similarity 95.3%; Pred. No. 0.021;
Matches 41; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTACTATAGGCGCGGTGTCGACGGCCCGGGCTGGTATGA 43
Db 107 CTCACATAGGCGCGGTGTCGACGGCCCGGGCTGGTCTGA 149

RESULT 3
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match          3.8%; Score 39.6; DB 1; Length 7218;
Best Local Similarity 5.0%; Pred. No. 0.035;
Matches 15; Conservative 163; Mismatches 122; Indels 0; Gaps 0;

QY 561 TGGCACCCGATGGATGTCCTCCCACTCGGAACCTGTCTCCAACTCTGCATTTTCGGTTCAGAGC 620
Db 1040 TGGCTGCAGTTCGAGGGAGCTTCGGATYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1099
QY 621 AGTACTGCATTTTGGCCATTGCTCGCTGCACGAGACTCGTTACTGTTCGGCGCGGAACC 680
Db 1100 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1159
QY 681 GGACCTCCCTCTGCTGCTGCCAACTGCCAACCCAGACCTGGCATGGCTGGTGGCGTCAT 740
Db 1160 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1219
QY 741 CTAATTCTCTATCAAGAATCACCTCCCATCTACTCCGCCCTGTGCGGCGACCTCGTGC 800
Db 1220 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1279
QY 801 CCTCGAATCCACGGGCACACGGTGCAGGCACACCCAGCCATCAAAATACGGCCTCTTT 860
Db 1280 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1339

RESULT 4
US-09-323-195A-1
; Sequence 1, Application US/09323195A
; Patent No. 6462257
; GENERAL INFORMATION:
; APPLICANT: Pullman, Gerald
; APPLICANT: Cairney, John
; APPLICANT: Ferrera, Ranjan
; TITLE OF INVENTION: VICILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND
; TITLE OF INVENTION: METHODS OF USING THE SAME
; FILE REFERENCE: IPST0009
; CURRENT APPLICATION NUMBER: US/09/323,195A
; CURRENT FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 341
; TYPE: DNA
; ORGANISM: Pinus taeda
US-09-323-195A-1

Query Match          3.8%; Score 39; DB 4; Length 341;
Best Local Similarity 89.4%; Pred. No. 0.011;
Matches 42; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 4 ACTATAGGCGACCGGTGGTCGACGGCCCGGGCTGGTATGAAGGTGGG 50
Db 1 ACTATAGGCGACCGGTGGTCGACGGCCCGGGCTGGTAAAGTGTGTG 47

RESULT 5
US-09-424-283-7
; Sequence 7, Application US/09424283
; Patent No. 6437219
; GENERAL INFORMATION:
; APPLICANT: Grimes, et al.
; TITLE OF INVENTION: Sucrose binding proteins
; FILE REFERENCE: 4630-50206
; CURRENT APPLICATION NUMBER: US/09/424,283
; CURRENT FILING DATE: 1999-11-19
```

; PRIOR APPLICATION NUMBER: PCT/US98/10465
; PRIOR FILING DATE: 1998-05-21
; PRIOR APPLICATION NUMBER: US 60/047,568
; PRIOR FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 4526
; TYPE: DNA
; ORGANISM: Glycine max
US-09-424-283-7

Query Match 3.8%; Score 39; DB 4; Length 4526;
Best Local Similarity 100.0%; Pred. No. 0.042;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTACTATAGGCACGGTGTGTCAGCGCCCGGGCTGGTA 40
|||||
Db 97 TTACTATAGGCACGGTGTGTCAGCGCCCGGGCTGGTA 135

RESULT 6

US-09-377-648-4/c
; Sequence 4, Application US/09377648
; Patent No. 6225529
; GENERAL INFORMATION:
; APPLICANT: Lappegard, Kathryn
; APPLICANT: Martino-Catt, Susan
; TITLE OF INVENTION: Seed-Preferred Promoters
; FILE REFERENCE: 0869
; CURRENT APPLICATION NUMBER: US/09/377,648
; CURRENT FILING DATE: 1999-08-19
; EARLIER APPLICATION NUMBER: US 60/097,233
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 921
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)...(922)
US-09-377-648-4

Query Match 3.7%; Score 38.4; DB 4; Length 921;
Best Local Similarity 97.5%; Pred. No. 0.028;
Matches 39; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCACGGTGTGTCAGCGCCCGGGCTGGTA 40
|||
Db 912 CTCACTATAGGCACGGTGTGTCAGCGCCCGGGCTGGTA 873

RESULT 7

US-09-570-367C-1
; Sequence 1, Application US/09570367C
; Patent No. 6338851
; GENERAL INFORMATION:
; APPLICANT: Gorczynski, Reginald M.
; TITLE OF INVENTION: Methods and Compositions for Immunomodulation
; FILE REFERENCE: 9579-21
; CURRENT APPLICATION NUMBER: US/09/570,367C
; CURRENT FILING DATE: 2000-05-05
; PRIOR APPLICATION NUMBER: US 60/064,764
; PRIOR FILING DATE: 1997-11-07
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 2791
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-570-367C-1

Query Match 3.7%; Score 38.4; DB 4; Length 2791;
Best Local Similarity 87.5%; Pred. No. 0.051;
Matches 42; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ACTATAGGCACGGTGTGTCAGCGCCCGGGCTGTGTAAGGTTGGGA 51
|||||
Db 1 ACTATAGGCACGGTGTGTCAGCGCCCGGGCTGTGTAAGGTTGGGA 48

RESULT 8

US-09-276-599-13/c
; Sequence 13, Application US/09276599
; Patent No. 6380459
; GENERAL INFORMATION:
; APPLICANT: Perera, J. Ranjan
; APPLICANT: Rice, Stephen J.
; TITLE OF INVENTION: Composition and methods for the
; TITLE OF INVENTION: modification of gene expression
; FILE REFERENCE: 11000.1036
; CURRENT APPLICATION NUMBER: US/09/276,599
; CURRENT FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 336
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-276-599-13

Query Match 3.7%; Score 38; DB 4; Length 336;
Best Local Similarity 100.0%; Pred. No. 0.022;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTACTATAGGCACGGTGTGTCAGCGCCCGGGCTGGT 39
|||||
Db 334 TTACTATAGGCACGGTGTGTCAGCGCCCGGGCTGGT 297

RESULT 9

US-09-276-599-14/c
; Sequence 14, Application US/09276599
; Patent No. 6380459
; GENERAL INFORMATION:
; APPLICANT: Perera, J. Ranjan
; APPLICANT: Rice, Stephen J.
; TITLE OF INVENTION: Composition and methods for the
; TITLE OF INVENTION: modification of gene expression
; FILE REFERENCE: 11000.1036
; CURRENT APPLICATION NUMBER: US/09/276,599
; CURRENT FILING DATE: 1999-03-25
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 14
; LENGTH: 763
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-276-599-14

Query Match 3.7%; Score 38; DB 4; Length 763;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 38; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTACTATAGGCACGGTGTGTCAGCGCCCGGGCTGGT 39
|||||
Db 761 TTACTATAGGCACGGTGTGTCAGCGCCCGGGCTGGT 724

RESULT 10

US-09-323-195A-5
; Sequence 5, Application US/09323195A
; Patent No. 6462257
; GENERAL INFORMATION:

APPLICANT: Pullman, Gerald
APPLICANT: Cairney, John
APPLICANT: Ferreira, Ranjan
TITLE OF INVENTION: VICILIN-LIKE SEED STORAGE PROTEIN GENE PROMOTER AND
TITLE OF INVENTION: METHODS OF USING THE SAME
FILE REFERENCE: IFST0009
CURRENT APPLICATION NUMBER: US/09/323,195A
CURRENT FILING DATE: 1999-06-01
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 565
TYPE: DNA
ORGANISM: Pinus taeda
FEATURE:
OTHER INFORMATION: n at 489 is a, c, g, or t
OTHER INFORMATION: n at 503 is a, c, g, or t
OTHER INFORMATION: n at 504 is a, c, g, or t
OTHER INFORMATION: n at 522 is a, c, g, or t
OTHER INFORMATION: n at 533 is a, c, g, or t
OTHER INFORMATION: n at 543 is a, c, g, or t
OTHER INFORMATION: n at 549 is a, c, g, or t
OTHER INFORMATION: n at 564 is a, c, g, or t
US-09-323-195A-5

Query Match 3.7%; Score 37.8; DB 4; Length 565;
Best Local Similarity 95.1%; Pred. No. 0.033;
Matches 39; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4 ACTATAGGCACGCGTGGTCGACGCCCGCGGCTGGTATGAA 44
|||||
Db 1 ACTATAGGCACGCGTGGTCGACGCCCGCGGCTGGTAAAAA 41

RESULT 11
US-09-424-283-5/C
Sequence 5, Application US/09424283
Patent No. 6437219
GENERAL INFORMATION:
APPLICANT: Grimes, et al.
TITLE OF INVENTION: Sucrose binding proteins
FILE REFERENCE: 4630-50206
CURRENT APPLICATION NUMBER: US/09/424,283
CURRENT FILING DATE: 1999-11-19
PRIOR APPLICATION NUMBER: PCT/US98/10465
PRIOR FILING DATE: 1998-05-21
PRIOR APPLICATION NUMBER: US 60/047,568
PRIOR FILING DATE: 1997-05-22
NUMBER OF SEQ ID NOS: 15
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 1924
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: CDS
LOCATION: (119)..(1588)
US-09-424-283-5

Query Match 3.7%; Score 37.8; DB 4; Length 1924;
Best Local Similarity 95.1%; Pred. No. 0.065;
Matches 39; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCACGCGTGGTCGACGCCCGCGGCTGGTAT 41
|||||
Db 1803 CTCACATAGGCACGCGTGGTCGACGCCCGCGGCTGGTTT 1763

RESULT 12
US-08-913-014A-18
Sequence 18, Application US/08913014A
Patent No. 6235878
GENERAL INFORMATION:

APPLICANT: Nishi, Kazunori
APPLICANT: Hikichi, Yukiko
APPLICANT: Shintani, Yasushi
TITLE OF INVENTION: NOVEL FAS LIGAND-LIKE PROTEIN, ITS
TITLE OF INVENTION: PRODUCTION AND USE
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: David G. Conlin, Esq.
ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & USHMAN, LLP
STREET: 130 Water Street
CITY: Boston,
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,014A
FILING DATE: 04-SEP-1997
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/02480
FILING DATE: July 17, 1997
ATTORNEY/AGENT INFORMATION:
NAME: David G. Conlin
REGISTRATION NUMBER: 27,026
REFERENCE/DOCKET NUMBER: 342/47694
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-523-3400
TELEFAX: 617-523-6440
TELEX:

INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 48
TYPE: Nucleic acid
STRANDEDNESS: Single
TOPOLOGY: Linear
MOLECULE TYPE: Synthetic DNA
US-08-913-014A-18

Query Match 3.6%; Score 37.4; DB 4; Length 48;
Best Local Similarity 97.4%; Pred. No. 0.012;
Matches 38; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTTACTATAGGCACGCGTGGTCGACGCCCGCGGCTGGT 39
|||||
Db 10 CTCACATAGGCACGCGTGGTCGACGCCCGCGGCTGGT 48

RESULT 13
US-09-570-367C-1/C
Sequence 1, Application US/09570367C
Patent No. 6338851
GENERAL INFORMATION:
APPLICANT: Gorczynski, Reginald M.
TITLE OF INVENTION: Methods and Compositions for Immunomodulation
FILE REFERENCE: 9579-21
CURRENT APPLICATION NUMBER: US/09/570,367C
CURRENT FILING DATE: 2000-05-05
PRIOR APPLICATION NUMBER: US 60/064,764
PRIOR FILING DATE: 1997-11-07
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 2791
TYPE: DNA
ORGANISM: Mus musculus
US-09-570-367C-1

Query Match 3.6%; Score 37.4; DB 4; Length 2791;

Best Local Similarity 97.4%; Pred. No. 0.11;
Matches 38; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 ACTATAGGCGACGCGTGTGCGACGGCGCGGCTGGTGATG 42
|||||
Db 2791 ACTATAGGCGACGCGTGTGCGACGGCGCGGCTGGTGCTG 2753

RESULT 14

US-09-545-814-28/c
; Sequence 28, Application US/09545814
; Patent No. 6416977
; GENERAL INFORMATION:
; APPLICANT: Becher, Anna M.
; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
; FILE REFERENCE: FC-5-C1
; CURRENT APPLICATION NUMBER: US/09/545,814
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,833
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 28
; LENGTH: 1478
; TYPE: DNA
; ORGANISM: Tagged Ctenocephalides felis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1467)
US-09-545-814-28

Query Match 3.6%; Score 37.2; DB 4; Length 1478;

Best Local Similarity 47.1%; Pred. No. 0.087;
Matches 114; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 66 ATATCTGCTGAGAGATAACATCACAAATTCACAGCTTCGATGTCACACAGCCTACAGC 125
|||
Db 456 AGACATTACGTGTAGCATCTAATATTCGCAAGCTCTGGTACATGATATCCCTCTCG 397
QY 126 CATTCCGAGCAGCGGCTTAGGACGGCGTCAATAACTTGGAGGAACTGCCAAGAAATGTG 185
|||
Db 396 CAGCCTAAATTTGGCAATCGGGACGCCATGGTGATCTCCCACTTTTTCCTTTGTTGT 337
QY 186 GATTACAACAGTTCTATCTGAGCGAGCTTAACAGCAGCTTGTATAGTCATTTTATC 245
|||
Db 336 GAAAGCCCTTCGCAATCTTGGCAAGTAATAAATTTGCTTTCTCGGAAAACTGCC 277
QY 246 TCCCCAACCGGCATTAGTAGCAGGCGCCCAACGCAATTCAAATTCGATGTAATTTGTCTAC 305
|||
Db 276 ACCTCTATCAGCAGCACCAGGATACTCCCAATCCAAGTCGAAACCGTCGAACTTTGTATTC 217
QY 306 TT 307
Db 216 GT 215

RESULT 15

US-09-545-814-30
; Sequence 30, Application US/09545814
; Patent No. 6416977
; GENERAL INFORMATION:
; APPLICANT: Becher, Anna M.
; TITLE OF INVENTION: FLEA CHITINASE NUCLEIC ACID MOLECULES, PROTEINS AND
; FILE REFERENCE: FC-5-C1
; CURRENT APPLICATION NUMBER: US/09/545,814
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: 60/128,833
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30

; LENGTH: 1478

; TYPE: DNA
; ORGANISM: Tagged Ctenocephalides felis
US-09-545-814-30

Query Match 3.6%; Score 37.2; DB 4; Length 1478;

Best Local Similarity 47.1%; Pred. No. 0.087;
Matches 114; Conservative 0; Mismatches 128; Indels 0; Gaps 0;

QY 66 ATATCTGCTGAGAGATAACATCACAAATTCACAGCTTCGATGTCACACAGCCTACAGC 125
|||
Db 1023 AGACATTACGTGTAGCATCTAATAATTCGCAAGCTCTGGTACATGATATCCCTCTCG 1082
QY 126 CATTCCGAGCAGCGGCTTAGGACGGCGTCAATAACTTGGAGGAACTGCCAAGAATGTG 185
|||
Db 1083 CAGCCTAAATTTGGCAATCGGGACGCCATGGTGATCTCCCACTTTTTCCTTTGTT 1142
QY 186 GATTACAACAGTTCTATCTGAGCGAGCTTAAACGCGAGCTTTGTATAGTCATTTTATC 245
|||
Db 1143 GAAAGCCCTTCGCAATCTTGGCAAGTAATAAATTTGCTTTTGTGCGAAAACTGCC 1202
QY 246 TCCCCAACCGGCATTAGTAGCAGGCGCCCAACGCAATTCAAATTCGATGTAATTTGTCTAC 305
|||
Db 1203 ACCTCTATCAGCAGCACCAGGATACTCCCAATCCAAGTCGAAACCGTCGAACTTTGTATTC 1262
QY 306 TT 307
Db 1263 GT 1264

Search completed: December 28, 2002, 00:01:55

Job time : 63 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 28, 2002, 00:00:48 ; Search time 1964 Seconds
(without alignments)
8493.560 Million cell updates/sec

Title: US-09-945-376-3

Perfect score: 1030

Sequence: 1 cttactataggcgcgctg.....tgtatatatacaaggtctgg 1030

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: gb_gss:*

18: em_gss_hum:*

19: em_gss_inv:*

20: em_gss_pln:*

21: em_gss_vrt:*

22: em_gss_fun:*

23: em_gss_mam:*

24: em_gss_mus:*

25: em_gss_other:*

26: em_gss_pro:*

27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	45	4.4	388	17 BH747511	BH747511 SALK_0180
c 2	43.6	4.2	469	17 BH251364	BH251364 SALK_0114
c 3	43.4	4.2	111	17 BH617520	BH617520 SALK_0372
c 4	42.8	4.2	124	17 BH613783	BH613783 SALK_0349
c 5	42.8	4.2	384	17 BH250931	BH250931 SALK_0107
c 6	42.8	4.2	445	17 BH634101	BH634101 SALK_0447

c 7	42.6	4.1	440	17 BH748887	BH748887 SALK_0469
c 8	42.2	4.1	432	17 BH610787	BH610787 SALK_0180
c 9	42.2	4.1	500	17 BH214178	BH214178 SALK_0103
c 10	42.2	4.1	1101	17 CNS00LGU	AL068402 Drosophila
c 11	42	4.1	129	17 BH168812	BH168812 SALK_0002
c 12	42	4.1	220	17 BH251572	BH251572 SALK_0118
c 13	42	4.1	452	17 BH212774	BH212774 SALK_0080
c 14	41.8	4.1	416	17 BH749138	BH749138 SALK_0473
c 15	41.6	4.0	457	17 BH172739	BH172739 SALK_0061
c 16	41.6	4.0	471	17 BH251074	BH251074 SALK_0109
c 17	41.6	4.0	502	17 BH253202	BH253202 SALK_0146
c 18	41.4	4.0	120	17 BH750440	BH750440 SALK_0408
c 19	41.4	4.0	430	17 BH612486	BH612486 SALK_0328
c 20	41.4	4.0	445	17 BH612500	BH612500 SALK_0328
c 21	41.4	4.0	456	17 BH617406	BH617406 SALK_0364
c 22	41.4	4.0	456	17 BH633739	BH633739 SALK_0431
c 23	41.2	4.0	404	17 BH617295	BH617295 SALK_0362
c 24	41.2	4.0	465	17 BH170845	BH170845 SALK_0031
c 25	41.2	4.0	469	17 BH749118	BH749118 SALK_0473
c 26	41	4.0	235	17 BH251060	BH251060 SALK_0109
c 27	41	4.0	248	17 BH254262	BH254262 SALK_0162
c 28	41	4.0	455	17 BH171573	BH171573 SALK_0045
c 29	40.8	4.0	433	17 BH610810	BH610810 SALK_0180
c 30	40.8	4.0	451	17 BH254421	BH254421 SALK_0155
c 31	40.8	4.0	477	17 BH170269	BH170269 SALK_0026
c 32	40.6	3.9	105	17 BH752334	BH752334 SALK_0140
c 33	40.6	3.9	121	17 BH616998	BH616998 SALK_0358
c 34	40.6	3.9	128	17 BH752369	BH752369 SALK_0184
c 35	40.6	3.9	218	17 BH170998	BH170998 SALK_0036
c 36	40.6	3.9	266	17 BH611050	BH611050 SALK_0237
c 37	40.6	3.9	454	17 BH611042	BH611042 SALK_0297
c 38	40.6	3.9	456	17 BH252514	BH252514 SALK_0135
c 39	40.6	3.9	492	17 BH611714	BH611714 SALK_0315
c 40	40.4	3.9	438	17 BH251902	BH251902 SALK_0123
c 41	40.4	3.9	460	17 BH188842	BH188842 036_P_10-
c 42	40.4	3.9	460	17 CNS07SRN	AL625781 t3 end of
c 43	40.4	3.9	470	17 BH172612	BH172612 SALK_0059
c 44	40.2	3.9	114	17 BH250920	BH250920 SALK_0107
c 45	40.2	3.9	116	17 BH251621	BH251621 SALK_0118

ALIGNMENTS

RESULT 1
BH747511/c 388 bp DNA linear GSS 27-FEB-2002
LOCUS SALK_018023.45.85.x Arabidopsis thaliana TDNA insertion lines
DEFINITION Arabidopsis thaliana genomic clone SALK_018023.45.85.x, DNA sequence.

ACCESSION BH747511
VERSION BH747511.1 GI:18960626
KEYWORDS GSS.

SOURCE thale cress.
ORGANISM Arabidopsis thaliana

REFERENCE 1 (bases 1 to 388)
AUTHORS Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab ,C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednisi,L., Shinn,P., Zimmerman,J. and Ecker,J.R.
A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome

Unpublished (2001)
Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (STGnAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: eckeresalk.edu

This is single pass sequence recovered from the left border of

RESULT 4
BH613783/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
QY
Db
QY
Db
RESULT 6
BH634101/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
QY
Db
QY
Db
RESULT 5
BH250931/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
QY
Db
QY
Db
RESULT 3
BH613783/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
QY
Db
QY
Db
RESULT 2
BH613783/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
QY
Db
QY
Db
RESULT 1
BH613783/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
QY
Db
QY
Db
RESULT 0
BH613783/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN
BASE COUNT
ORIGIN
Query Match
Best Local Similarity
Matches
QY
Db
QY
Db

TITLE Zimmermann, J. and Ecker, J. R.
A Sequence-Indexed Library of Insertion Mutations in the
Arabidopsis Genome

JOURNAL Unpublished (2001)

COMMENT Contact: Joseph R. Ecker
Salk Institute Genomic Analysis Laboratory (SIGNAL)
The Salk Institute for Biological Studies
10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
Tel: 858 453 4100 x1752
Fax: 858 558 6379
Email: ecker@salk.edu
This is single pass sequence recovered from the left border
TDNA.

```

Class: TDNA tagged.
Location/Qualifiers
1..129
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_000216"
/clone_lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on Arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"
25 a 38 c 28 g 38 t
BASE COUNT
ORIGIN

```

	Query Match	4.1%	Score 42;	DB 17;	Length 129;
	Best Local Similarity	64.3%;	Pred. No. 0.054;		
	Matches 63;	Conservative 0;	Mismatches 35;	Indels 0;	Gaps 0;
QY	1	CTTACTATAGGCGACGGCTGTGCACGCCGCCGGCTGGTATGAAGTTGGGAACCTCACTG	60		
Db	123	CTCATATAGGCGACGGCTGTGCACGCCGCCGGCTGCAATTCCGTTGAAGAATCAGAG	64		
QY	61	GATGCATATACTGCTGAGAGATACATCAATTCACA	98		
Db	63	AGAAACAGAACCAAGCGAGAGACGTTGGTGTTTAACA	26		

RESULT 12	
BH251572/c	
LOCUS	BH251572
DEFINITION	SALK_011803 Arabidopsis thaliana TDNA insertion lines Arabidopsis thaliana genomic clone SALK_011803, DNA sequence.
ACCESSION	BH251572.1
VERSION	GI:17138550
KEYWORDS	GSS.
SOURCE	thale cress.
ORGANISM	Arabidopsis thaliana
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE	1 (bases 1 to 220)
AUTHORS	Alonso,J.M., Leisse,T.J., Barajas,P., Chen,H., Cheuk,R., Gadrinab , C., Jeske,A., Karnes,M., Kim,C.J., Parker,H., Prednis,L., Shinn,P., Zimmerman,J. and Ecker,J.R.

TITLE 7 Zimmermann, J. and Eckert, J. K.
 A Sequence-Indexed Library of Insertion Mutations in the
 Arabidopsis Genome
 JOURNAL Unpublished (2001)
 COMMENT Contact: Joseph R. Ecker
 Salk Institute Genomic Analysis Laboratory (STGnAL)
 The Salk Institute for Biological Studies
 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA
 Tel: 858 453 4100 x1752
 Fax: 858 558 6379
 Email: ecker@salk.edu
 This is single pass sequence recovered from the left border of
 TDNA.
 Class: TDNA tagged.

FEATURES
source

Location/Qualifiers
1 220

```

1. .220
location/Qualifiers
/organism="Arabidopsis thaliana"
/strain="Columbia 0"
/db_xref="taxon:3702"
/clone="SALK_011803"
/clone.lib="Arabidopsis thaliana TDNA insertion lines"
/note="PCR was performed on arabidopsis thaliana lines
each of which contains one or more TDNA insertion
elements. The resultant fragment for each line was
directly sequenced to determine the genomic sequence at
the site of insertion. Details of the protocols used can
be found at http://signal.salk.edu/tdna\_protocols.html"

```

```

BASE COUNT      44 a      35 c      48 g      93 t
ORIGIN
Query Match      4.1%; Score 42; DB 17; Length 220;
Best Local Similarity 73.0%; Pred. No. 0.067;
Matches 54; Conservative 0; Mismatches 20; Indels 0; Gaps 0;

QY      1  CTTACTATAGGCACGCGTGGTCGACGCCCGCGGCTGCTGTATGAAGGTGGGAACCTCACTG 60
      || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      214  CTCACATATAGGCACGCGTGGTCGACGCCCGCGGCTCAATTCAAGTCTTAACACAAT 155

QY      61  GATGCATATATGCG 74
      | | | | |
Db      154  TTTTAAATTAGTGC 141

```

RESULT 13	BH212774.c
LOCUS	BH212774
DEFINITION	BH212774
ACCESSION	BH212774.1
VERSION	GSS.
KEYWORDS	Keywords
SOURCE	Source
ORGANISM	Organism
	Arabidopsis thaliana
	thale cress.
	BH212774.1
	GI:16394148
	DNA
	linear
	GSS 24-OCT-2001
	SALK_08030 Arabidopsis thaliana TDNA insertion lines Arabidopsis
	thaliana genomic clone SALK_08030, DNA sequence.

REFERENCE	1. (bases 1 to 452)
AUTHORS	Alonso, J.M., Leisse, T.J., Barajas, P., Chen, H., Cheuk, R., Gadrinab, C., Jeske, A., Karnes, M., Kim, C.J., Parker, H., Prednis, L., Shinn, P., Zimmerman, J. and Ecker, J.R.
TITLE	A Sequence-Indexed Library of Insertion Mutations in the Arabidopsis Genome
JOURNAL	Unpublished (2001)
COMMENT	Contact: Joseph R. Ecker Salk Institute Genomic Analysis Laboratory (SIGnAL) The Salk Institute for Biological Studies 10010 N. Torrey Pines Road, La Jolla, CA 92037, USA Tel: 858 453 4100 x1752 Fax: 858 558 6379 Email: ecker@salk.edu This is single pass sequence recovered from the left border of TDNA.

FEATURES	SOURCE
----------	--------

Location/Qualifiers
1. .452

BASE COUNT

```

114 a 113 c 111 q 93 t

```

	Query Match	4.1%	Score 41.8;	DB 17;	Length 416;
	Best Local Similarity	80.3%;	Pred. No. 0.099;		
	Matches 49;	Conservative	0;	Mismatches 12;	Indels 0; Gaps 0;
Qy	1	CTTACTATAGGGCACGCGTGGTTCGACGCCGGGGCTGGTATGAAGTGGGAACCTCACTG	60		
Db	156	CTCACTATAGGGCACGCGTGGTTCGACGCCGGGGCTGGTATGAAGTGGGAACCTCACTG	97		

QY 65 CATATATCGTGGAGAGATAACATCACAAATTCAGAAGCTTCGATGTCATCTACA 116

DG 192 GACGTAACCTTTGAGGGTTATCTCTCCTTACACATTTTATGAGACACCA 141

Search completed: December 28, 2002, 01:25:34
Job time : 1971 secs

